Computational Studies of Transmembrane Helix Insertion and Association

Alan Chetwynd



Structural Bioinformatics & Computational Biochemistry

Department of Biochemistry

and

Keble College, Oxford

Trinity 2011

A thesis submitted in partial fulfilment of the

requirements for the degree of Doctor of Philosophy

at the University of Oxford

Abstract

Computational Studies of Transmembrane Helix Insertion and Association

Alan Chetwynd

Structural Bioinformatics & Computational Biochemistry Unit,
Department of Biochemistry
and
Keble College, Oxford

Submitting for the degree of Doctor of Philosophy, Trinity 2011

Membrane proteins perform a variety of functions essential for the viability of the cell, including transport and signalling across the membrane. Most membrane proteins are formed from bundles of transmembrane helices. In this thesis molecular dynamics simulations have been used to investigate helix insertion into bilayers and helix association within bilayers. The potentials of mean force for the insertion of helices derived from the cystic fibrosis transmembrane conductance regulator into lipid bilayers were calculated using coarse-grained molecular dynamics simulations. The results showed that the insertion free energy increased with helix length and bilayer hydrophobic width. The insertion free energies obtained were significantly larger than comparable quantities obtained from translocon-mediated insertion experiments, consistent with a variety of previous studies. The implications of this observation for the interpretation of *in vivo* translocon-mediated insertion experiments, and the function of the translocon, are discussed.

Coarse-grained and atomistic molecular dynamics simulations of the transmembrane region of the receptor tyrosine kinase EphA1 suggested that the transmembrane helix dimer was most stable when interacting via the glycine zipper motif, in agreement with a structure obtained by NMR spectroscopy. Coarse-grained simulations of the transmembrane region of EphA2 suggested that the dimer has two stable orientations, interacting via a glycine zipper or a heptad motif. Both structures showed right-handed dimers, although an NMR structure of the transmembrane region of EphA2 shows a left-handed dimer interacting via the heptad motif. Both structures obtained from coarse-grained simulations proved unstable when simulated at an atomistic level of detail.

The potentials of mean force for dissociating the EphA1 and EphA2 dimers were calculated using coarse-grained molecular dynamics calculations. Convergence of the detailed structure of the profiles was not conclusively shown, although association free energies calculated from the profiles were consistent over a variety of simulation times. The association free energies were slightly larger than experimental values obtained for comparable systems, but consistent with similar computational calculations previously reported. However, direct comparisons are difficult owing to the influence of environmental factors on reported association free energies. The potential of mean force profiles showed that the interaction via the glycine zipper motif for EphA1 was significantly more stable than any other conformation. For EphA2 the potential of mean force profiles suggested that interaction via the glycine zipper and heptad motifs both provided stable or metastable conformations, with the interaction via the glycine zipper motif probably at least as stable as that via the heptad motif.

Acknowledgements

A lot of people have helped with this thesis. Firstly, I'd like to thank Mark. I couldn't have asked for a more supportive or helpful supervisor. There may have been a few occasions when running took a slightly higher priority than it ought to have done, so I'd particularly like to thank Mark for showing great restraint and pointing this out to me only once a month or so.

Secondly, I'd like to thank everyone at SBCB. Particular mention should go to Kathryn, who was enormously helpful when I first joined the lab; Craig, who has happily answered all my idiotic questions despite my inability to use Google to solve my problems; and Phyllis, who has had to endure countless daft questions on everything from computers to midlife crises. Many others have been extremely helpful, including (but not limited to) Ben, Chesney, Rouse, Ranjit, Oli, Goosey, PhilF, Tyler, Jerome, Manfred, Danny, Mickael, and Tim.

And lastly, I'd like to thank my friends and family, who have helped keep me (reasonably) sane all these years.

Oxford, September 2011

Word Count: 30,014

List of Publications Arising from this Thesis

A. Chetwynd, C. L. Wee, B. A. Hall, and M. S. P. Sansom, 2010. The Energetics of Transmembrane Helix Insertion into a Lipid Bilayer. *Biophys. J.* 99:2534-2540

List of Publications Arising from Related Work during my DPhil

A. Chetwynd, K. A. Scott, Y. Mokrab, and M. S. P. Sansom, 2008. CGDB: A Database of Membrane Protein-Lipid Interactions by Coarse-Grained MD Simulations. *Mol. Mem. Biol.* 25:662-669.

C. L. Wee, A. Chetwynd, and M. S. P. Sansom, 2011. Membrane Insertion of a Voltage Sensor Helix. *Biophys. J.* 100:410-419

B. A. Hall, A. Chetwynd, and M. S. P. Sansom, 2011. Exploring Peptide-Membrane Interactions with Coarse-Grained MD Simulations. *Biophys. J.* 100:1940-1948

List of Abbreviations

ADP adenosine-diphosphate

ATP adenosine-triphosphate

CG coarse-grained

cryo-EM cryo electron microscopy

DHPC 1,2-di-heptanoyl-*sn*-glycero-3-phosphocholine

DLPC 1,2-dilauroyl-*sn*-glycero-3-phosphatidyl-choline

DMPC 1,2-di-myristoyl-*sn*-glycero-3-phosphatidyl-choline

DOPC 1,2-di-oleoyl-*sn*-glycero-3-phosphatidyl-choline

DPD dissipative particle dynamics

DPPC 1,2-di-palmitoyl-*sn*-glycero-3-phosphatidyl-choline

ENM elastic network model

Eph erythropoietin-producing hepatocellular

ER endoplasmic reticulum

FRET fluorescence resonance energy transfer

GB generalized Born

GpA Glycophorin A

GPCR G-protein coupled receptor

GROMACS GROningen MAchine for Chemical Simulations

GTP Guanosine triphosphate

H-bonds hydrogen bonds

IF interfacial

Kv voltage-gated potassium (e.g. Kv channels)

l-bfgs limited memory Broyden-Fletcher-Goldfarb-Shanno quasi Newton minimiser

MC Monte Carlo

MD molecular dynamics

MM molecular mechanics

MSD mean-squared distance

NMR nuclear magnetic resonance

PBC periodic boundary conditions

PDB protein data bank

PME particle mesh Ewald

PMF potential of mean force

POPC 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidyl-choline

QM quantum mechanics, quantum mechanical (e.g. QM techniques)

RMSD root mean squared deviation

RNC ribosome nascent chain

RTK receptor tyrosine kinase

SAS solvent accessible surface

SD standard deviation

SRP signal recognition particle

TM transmembrane

VDW van der Waals

WHAM weighted histogram analysis method

List of Tables

2.1 Amino acid sidechain particle types for the CG model used in this thesis [1]. . 27

3.1 Different hydrophobicity scales 37
3.2 PMF calculations performed 42
3.3 Sequences of helices 44
3.4 Free energies obtained from PMF profiles 46
3.5 M8 Sequence 64
3.6 Results from the additive model for free energy changes 74
3.7 Results from the additive model for PMF depths 74
3.8 Estimated insertion free energies from additive model using IF free energies . 77

4.1 Simulated EphA1 and EphA2 sequences 89
4.2 Summary of 'ideal' EphA2 simulations. 91
4.3 Summary of 'ideal' EphA1 simulations. 102

5.1 Summary of PMF simulations 122
5.2 Association free energies from PMF calculations 147

List of Figures

1.1 A section of an Escherichia coli cell. . . . 3
1.2 Monotopic, bitopic and polytopic membrane proteins. . . . 4
1.3 Sample membrane proteins in lipid bilayers. . . . 5
1.4 Translocon crystal structure 1RH5. . . . 7
1.5 Insertion of TM helices by the translocon. . . . 9
1.6 Translocon-mediated insertion experiments. . . . 10
1.7 Approximate time and length scales accessible from different simulation methods. . . . 12

2.1 A DPPC molecule . . . 26

3.1 CFTR . . . 40
3.2 Helices of different lengths . . . 45
3.3 PMF profiles for M2-19, M2-29 and M2-39 . . . 45
3.4 Bilayer distortion . . . 47
3.5 M2-19 in an interfacial position . . . 47
3.6 Average bilayer widths . . . 48
3.7 Snapshots from M2-39 simulations . . . 48
3.8 Helix tilt angles from M2-19, M2-29 and M2-39 simulations . . . 49
3.9 M2-39 snapshots . . . 51
3.10 M2-29 PMF profile, unmanipulated . . . 52
3.11 M2-29 snapshots . . . 52
3.12 M2-29 PMF profiles, manipulated . . . 53
3.13 Model M2-29 free energy landscape . . . 54
3.14 PMF profiles for different bilayer widths . . . 55
3.15 CG lipids used . . . 56
3.16 Correlation of ΔG_{pmf} with bilayer widths . . . 57
3.17 Helix tilt angles for thinner bilayer simulations . . . 58
3.18 PMF profiles for M2-39 with different starting orientations . . . 60
3.19 Snapshot of M2-39_{90} . . . 60
3.20 Tilt angle of M2-39_{90} as a function of distance from the bilayer. . . . 61
3.21 Schematic of suggested energy landscape for M2-39 . . . 61
3.22 Frequency of displacement of helices from the restraint points . . . 62
3.23 PMF profile varying secondary structure restraints . . . 63
3.24 Snapshot of M2-$39_{partial}$ in a bilayer . . . 64
3.25 PMF profiles for M8 . . . 65
3.26 Helix tilt distributions for M8 . . . 66
3.27 Snapshots from M8 simulations . . . 67
3.28 PMFs for M8, manipulated . . . 68
3.29 Symmetrised M8 PMF profile . . . 69
3.30 PMF profiles indicating the effect of clamping . . . 71
3.31 PMF profiles showing convergence . . . 72
3.32 Comparing M2-39 PMF with experimental data . . . 78
3.33 A schematic for helix insertion . . . 81

4.1 Left and right handed dimers . . . 87
4.2 NMR structures of EphA1 and EphA2 . . . 88
4.3 Snapshots from EphA2 CG simulations . . . 92
4.4 Rotation angles explained . . . 93
4.5 Rotation angles for EphA2 simulations . . . 94
4.6 Rotation angle distribution for EphA2 simulations . . . 95
4.7 Snapshots from EphA2 simulations . . . 96
4.8 Particle-particle distances for EphA2 simulations . . . 97
4.9 Analysis of EphA2 atomistic simulations . . . 99
4.10 Snapshots from the EphA2 atomistic simulations starting with the glycine zipper motifs facing . . . 100
4.11 Snapshots from the EphA2 atomistic simulations starting with the heptad motifs facing . . . 101
4.12 Rotation of EphA1 helices . . . 102
4.13 Rotation angle distribution for EphA1 helices . . . 103
4.14 Histograms of the helix centre of mass separation in the five unrestrained 'ideal' EphA1 simulations. . . . 104
4.15 Helix centres of mass separation . . . 105
4.16 Snapshots from EphA1 simulations . . . 106
4.17 Particle-particle distances for EphA1 simulations . . . 107
4.18 EphA1 atomistic simulations analysis . . . 108
4.19 Snapshots from atomistic EphA1 simulations . . . 109
4.20 Comparison of CG EphA2 dimer structures with NMR structures . . . 112
4.21 Comparison of CG EphA2 dimer structures with NMR structures . . . 113

5.1 PMF simulations . . . 121
5.2 PMF for glycine-restrained EphA2 simulations . . . 123
5.3 Snapshots from EphA2 PMF . . . 124
5.4 Rotation angles per window for glycine-restrained EphA2 simulations . . . 125
5.5 Rotation angle distribution for glycine-restrained EphA2 simulations . . . 126
5.6 PMF for heptad-restrained EphA2 simulations . . . 127
5.7 Snapshots from heptad-restrained EphA2 simulations . . . 128
5.8 Rotation angles per window for heptad-restrained EphA2 PMF . . . 129
5.9 Rotation angle distribution for heptad-restrained EphA2 PMF . . . 130
5.10 PMF profiles for NMR structure EphA2 simulations . . . 131
5.11 Rotation angle distributions for NMR structure simulations . . . 133
5.12 Average rotation angle of EphA2 from NMR-derived helix simulations . . . 134
5.13 Snapshots of EphA2 from NMR-derived helix simulations with glycine zipper restrained . . . 134
5.14 Snapshots of EphA2 from NMR-derived helix simulations with heptad restrained . . . 135
5.15 Snapshots of EphA2 from NMR-derived helix simulations with Cαs restrained 135
5.16 EphA1 PMF . . . 136
5.17 EphA1 snapshots . . . 137
5.18 EphA1 rotation angles per window . . . 138
5.19 EphA1 rotation angle distribution . . . 139
5.20 Association free energies as a function of bin width . . . 140
5.21 Association free energies as a function of simulation time . . . 141
5.22 Forces acting on the helices . . . 142
5.23 PMFs calculated using different WHAMs . . . 143
5.24 EphA2 convergence . . . 143

Contents

Abstract i

Acknowledgements ii

List of Publications Arising from this Thesis iii

List of Abbreviations iv

List of Tables vi

List of Figures vii

1 Introduction 1
- 1.1 What is this Thesis about? . . . 1
- 1.2 A Brief Introduction to Membranes . . . 1
- 1.3 Membrane Proteins . . . 2
- 1.4 Helices as Subunits of Membrane Proteins . . . 5
- 1.5 Helical Membrane Protein Folding: The Two Stage Model . . . 6
  - 1.5.1 Stage One: Entry via the Translocon . . . 6
  - 1.5.2 Stage Two: Helix-Helix Aggregation . . . 10
- 1.6 Simulations of Membrane Proteins . . . 11
  - 1.6.1 Coarse-Grained Simulations . . . 13
- 1.7 Simulations of Helices . . . 14
- 1.8 Details of Work Presented in this Thesis . . . 15

2 Theory & Methods 17
- 2.1 Introduction . . . 17
- 2.2 Molecular Dynamics Simulations . . . 17
  - 2.2.1 The Born-Oppenheimer Approximation . . . 17
  - 2.2.2 Solving Newton's Equations . . . 18
  - 2.2.3 Force fields . . . 19
  - 2.2.4 Periodic Boundary Conditions . . . 21
  - 2.2.5 Energy Minimization . . . 22
  - 2.2.6 Temperature and Pressure Coupling . . . 23
  - 2.2.7 Efficient Electrostatics . . . 23
  - 2.2.8 Bond Constraints . . . 24
  - 2.2.9 Position Restraints . . . 24
- 2.3 Coarse-Grained Simulations . . . 24
  - 2.3.1 Coarse-Graining . . . 24
  - 2.3.2 The CG Lipid and Protein Model . . . 25
- 2.4 Free Energy Calculations . . . 27
  - 2.4.1 The Potential of Mean Force . . . 28
  - 2.4.2 Umbrella Sampling . . . 28
  - 2.4.3 Weighted Histogram Analysis Method (WHAM) . . . 29

2.4.4 Other Methods for Calculating Free Energies . . . 31
2.5 Details of Simulations Presented in this Thesis . . . 33

3 Helix-Bilayer PMF Calculations 35
3.1 Introduction . . . 35
3.1.1 CFTR . . . 38
3.2 Methods . . . 39
3.2.1 Theory . . . 39
3.2.2 Relationship between PMF and experimental results . . . 39
3.2.3 PMF Calculations . . . 42
3.2.4 Simulation Details . . . 42
3.3 Results . . . 44
3.3.1 Helix Length . . . 44
3.3.2 Bilayer Width . . . 55
3.3.3 Initial Orientation . . . 57
3.3.4 Folding State . . . 59
3.3.5 Helix Hydrophobicity . . . 64
3.4 Discussion . . . 70
3.4.1 Sampling . . . 70
3.4.2 Comparison of Results . . . 73
3.5 Conclusions . . . 83

4 Multiscale Simulations of Eph Receptor Transmembrane Helices 85
4.1 Introduction . . . 85
4.1.1 Receptor Tyrosine Kinases . . . 85
4.1.2 Eph Receptors . . . 85
4.1.3 EphA1 and EphA2 . . . 86
4.2 Methods . . . 89
4.2.1 Sequences . . . 89
4.2.2 Coarse-Grained Simulations . . . 90
4.2.3 Coarse Grained to Atomistic Conversion . . . 90
4.2.4 Analysis . . . 91
4.3 Results . . . 91
4.3.1 EphA2 . . . 91
4.3.2 EphA1 . . . 102
4.4 Discussion . . . 109
4.4.1 Dimerising Time . . . 109
4.4.2 EphA2 Interaction Modes . . . 110
4.4.3 EphA1 Interaction Modes . . . 111
4.5 Conclusions . . . 114

5 PMF Calculations of Eph Receptor Transmembrane Helices 115
5.1 Introduction . . . 115
5.2 Methods . . . 118
5.2.1 Theory . . . 118
5.2.2 Relationship between PMF and experimental results . . . 118
5.2.3 Simulation Details . . . 120
5.2.4 PMF Calculations . . . 121
5.3 Results . . . 122
5.3.1 EphA2 - Ideal Helix . . . 122
5.3.2 EphA2 - NMR structure . . . 129
5.3.3 EphA1 - Ideal Helix . . . 132
5.4 Sensitivity Analysis . . . 138
5.4.1 Calculating Equilibration Time . . . 138

5.4.2 Implementing WHAM . . . 140
5.4.3 Convergence . . . 141
5.4.4 Effect of Choice of Reaction Coordinate . . . 144
5.5 Discussion . . . 146
5.5.1 Comparison with Similar Systems . . . 146
5.5.2 EphA2: Two Interaction States? . . . 146
5.5.3 EphA2 vs EphA1 stability . . . 148
5.6 Conclusions . . . 148

6 Conclusions **150**
6.1 Summary . . . 150
6.2 Future Work . . . 152
6.3 Conclusions . . . 154

Chapter 1

Introduction

1.1 What is this Thesis about?

This thesis describes the use of molecular dynamics (MD) simulations to investigate processes related to the folding of α-helical membrane proteins. I have used coarse-grained (CG) MD simulations to calculate the potential of mean force (PMF) of helix insertion into a lipid bilayer, and the dimerisation of helices within a lipid bilayer. These two processes are representative of the two stages that are necessary for the folding of an α-helical membrane protein in a lipid bilayer. The PMF of transmembrane helix insertion may be compared with similar results from translocon-mediated insertion experiments to provide insights into possible mechanisms of *in vivo* helix insertion. CGMD and atomistic simulations are used to investigate the interaction modes of the transmembrane (TM) region of two receptor tyrosine kinases, EphA1 and EphA2. Both unrestrained simulations and PMF calculations suggest multiple possible interaction modes for EphA2, but only one stable interaction mode for EphA1. However, difficulties with sampling suggest that the PMF method is not accurate enough to confidently resolve interaction energy differences of $<$10 kJ/mol. To provide a context for this work, a discussion of membrane proteins, membrane protein folding and MD simulations is presented in this chapter.

1.2 A Brief Introduction to Membranes

For an organism to be viable, it must create a barrier to differentiate itself from its environment. Membranes consisting of lipids (including phospholipids, sphingolipids and sterols), proteins and carbohydrates provide this compartmentalisation for prokaryotes and eukaryotes. The basic unit of a membrane is a bilayer formed by phospholipids and sphin-

golipids in two layers, with their polar headgroups surrounding the non-polar central domain formed by the lipid chains. Embedded into the bilayer are membrane proteins. Such proteins are often densely packed in the membrane; the lipid concentration of membranes typically varies from 20% to 80% by dry weight [2]. Membranes are fluid structures, with proteins and lipids able to move laterally to facilitate multiple interactions [3].

Membranes are not simply passive barriers; they must be selectively permeable to specific molecules, and allow the transduction of signals into the cell. These transport and signalling processes are typically performed by membrane proteins, with a complex interdependence between many of these proteins and their lipid environment [4, 5]. An example of the importance of lipids in membranes is the increasing attention on the role of lipid rafts in membrane function. Lipid rafts are areas of the membrane enriched in cholesterol, giving them structural properties - such as a thicker hydrophobic width - which help differentiate them from other areas of the membrane [6, 7] and may help the membrane segregate its contents laterally.

An understanding of the structure and function of membranes needs to include an appreciation of the dynamic and crowded environment of the cell (see Figure 1.1 for a sample illustration). There are a massive variety of proteins performing a range of essential activities in any cell; the complexity and range of these interactions means that a reductionist approach - trying to understand a small part of the cell, often in isolation from its natural environment - has often been used to try to further our understanding of cellular function. Much of this thesis implicitly relies on a reductionist approach, trying to understand the factors that affect the interactions of helices within membranes. However, it should be appreciated that the pictures of individual helices within a small patch of a lipid bilayer often presented in this thesis are a huge simplification of the cellular environment.

1.3 Membrane Proteins

As mentioned above, membrane proteins play a vital role in many of the functions performed by a membrane. The importance of membrane proteins for cellular function can be seen in the fact that $\sim$25% of the human genome codes for membrane proteins [8].

Membrane proteins can be classified in a number of ways. A straightforward method of classification is by the secondary structure of their transmembrane (TM) region: α-helical membrane proteins have a TM region which consists of bundles of α-helices; β-barrel proteins have a TM region formed by β-sheets. The majority of membrane proteins are α-helical,

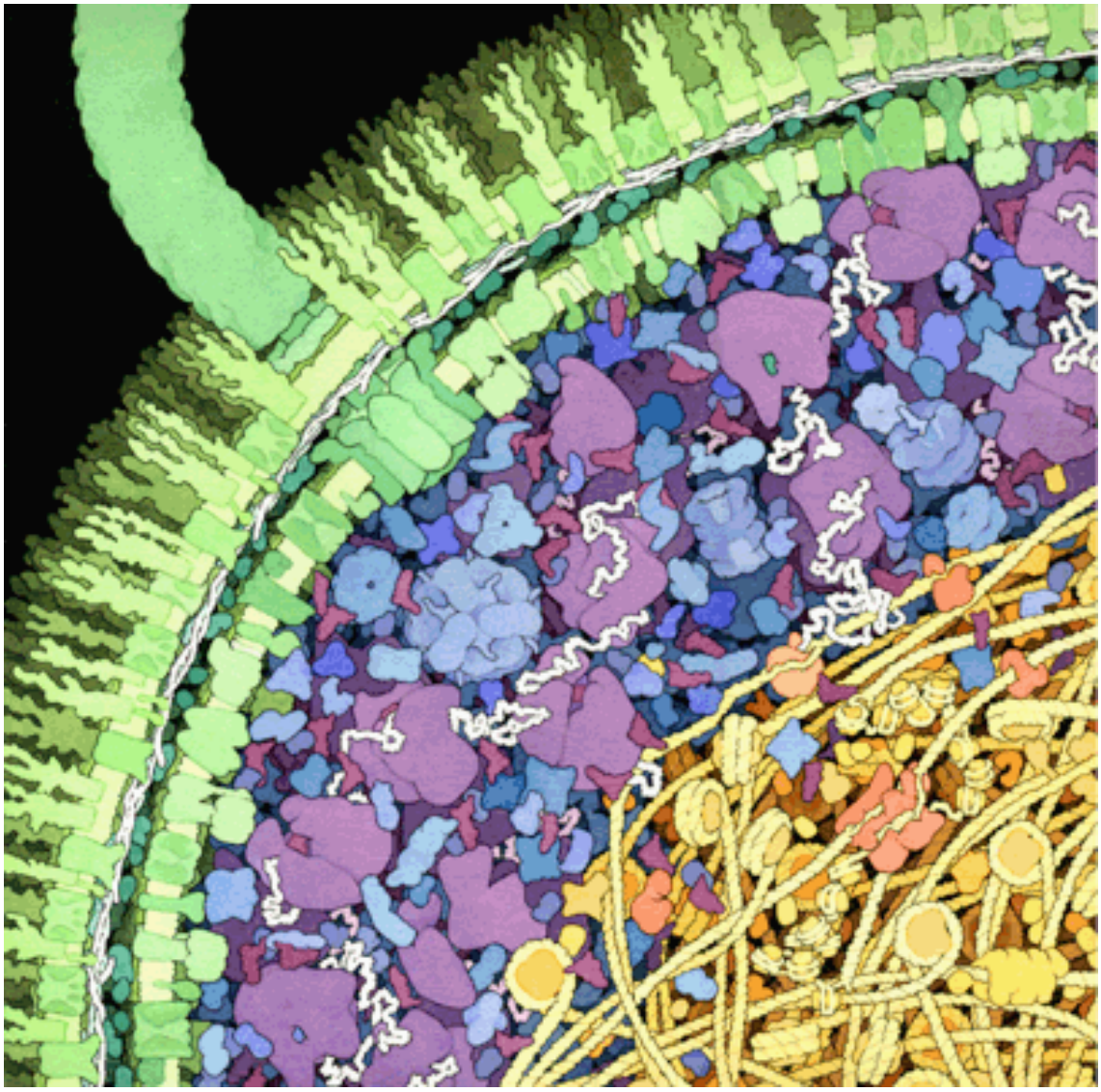


Figure 1.1: A section of an Escherichia coli cell. The cell wall, with two concentric membranes studded with transmembrane proteins, is shown in green. A large flagellar motor crosses the entire wall. The cytoplasmic area is shown in blue and purple. The large purple molecules are ribosomes and the small, L-shaped maroon molecules are tRNA, and the white strands are mRNA. Enzymes are shown in blue. The nucleoid region is shown in yellow and orange, with the long DNA circle shown in yellow, wrapped around HU protein (bacterial nucleosomes). DNA polymerase is shown in red-orange. Image courtesy of David Goodsell.

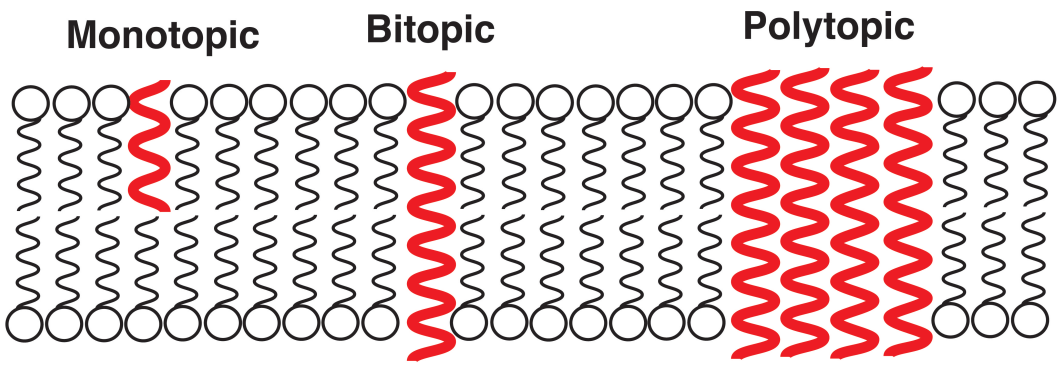


Figure 1.2: A structural method of characterising membrane proteins. Monotopic proteins (left) bind to one leaflet of the membrane. Bitopic membrane proteins (middle) span the membrane once. Polytopic membrane proteins (right) span the membrane multiple times.

although proteins in the outer membrane of Gram-negative bacteria are mostly formed from β-barrels. This thesis is solely concerned with α-helical proteins.

Membrane proteins can also be classified by their method of interaction with the membrane. Monotopic proteins bind to but do not fully span the membrane - examples include the cycloxygenase proteins (COX) [9]. Bitopic proteins, such as the receptor tyrosine kinase (RTK) family [10], span the membrane once. Polytopic proteins span the membrane multiple times; examples include the G-protein coupled receptor (GPCR) family [11]. See Figure 1.2 for an illustration.

The function of membrane proteins provides a third method of classification. For example, proteins which facilitate the movement of molecules across the membrane can be classified as either channels, pumps or transporters. Channels, such as aquaporins [12] or potassium channels [13], permit flow of a molecule down its concentration gradient. Pumps, such as the transmembrane ATPases [14], use energy from ATP hydrolysis to transport a molecule against its concentration gradient. Transporters, such as the major facilitator superfamily (MFS) transporters [15], use the transport of one molecule down its concentration gradient to achieve the transport of a second molecule against its concentration gradient.

Despite the importance of membrane proteins, relatively few high-resolution structures are available. At the time of writing, only 302 of the ~75,000 structures in the Protein Data Bank (PDB) represent unique membrane proteins (i.e. a membrane protein with a unique sequence; an up-to-date list can be found at http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). This is due to a number of problems: membrane proteins are difficult to express, owing to their dependence on the (limited volume) membrane environment; subsequent removal

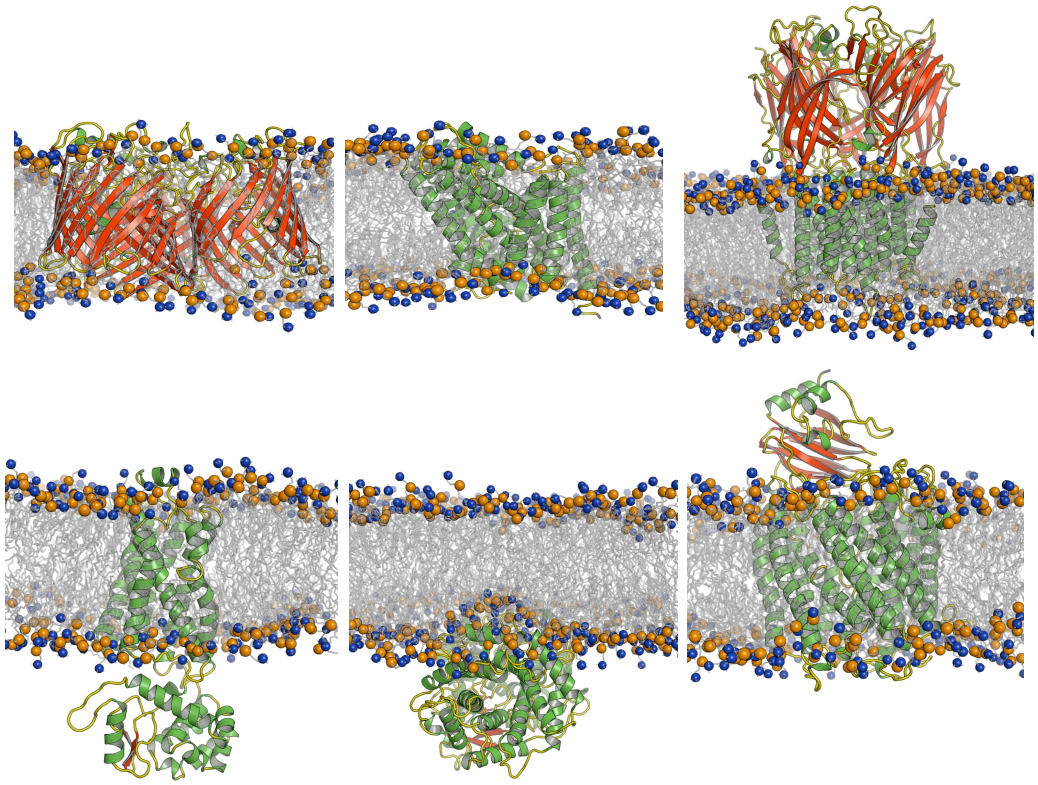


Figure 1.3: Examples of membrane proteins in lipid bilayers. Helices are in green, β-sheets in red and coil regions in yellow. Note the α-helical nature of the transmembrane region of most proteins. Image courtesy of Dr Phillip Stansfeld.

from the membrane is likely to compromise their structure; they are difficult to crystallise for X-ray crystallography; and tumbling of a protein/detergent assembly may lead to peak broadening for nuclear magnetic resonance (NMR) spectroscopy [16].

1.4 Helices as Subunits of Membrane Proteins

As discussed above, the transmembrane region of most membrane proteins consists of bundles of α-helices (see Figure 1.3). However, many membrane proteins also contain large intra- and/or extra-cellular domains, which may consist of combinations of α-helices, β-sheets and coil regions. Transmembrane helices are typically $\sim$12-25 residues in length, sufficient to span the hydrophobic region of the membrane, although defining the precise start and end residues of a given helix can be difficult. The length of a given helix is therefore often somewhat subjective, even when the structure of the protein from which it is derived is known. Helices within membrane proteins can adopt a variety of tilts and kinks to adapt to their environment.

1.5 Helical Membrane Protein Folding: The Two Stage Model

How membrane proteins enter the membrane is a topic of current interest [17, 18, 19]. The insertion of α-helical membrane proteins, with which this thesis is concerned, has been usefully described by the two stage model of Popot and Engelman [20], with some additions [21]. This model considers the insertion process as two thermodynamically independent stages: firstly, the helices insert into the membrane; secondly, they assemble into a functional protein.

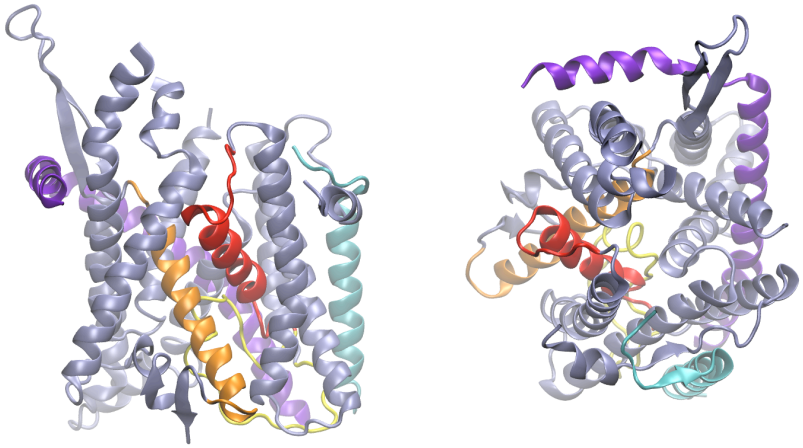
1.5.1 Stage One: Entry via the Translocon

The insertion of helices into the cell membrane is mediated by a highly conserved membrane protein known as Sec61 in eukaryotes, and SecY in bacteria. The Sec61/SecY proteins - both also known as the translocon - perform a number of complicated tasks. A translocon provides a pore for an unfolded protein; it decides whether this polypeptide chain should enter the membrane, or pass through into the cell; and it manages the orientation of the chain, all while maintaining a tight seal to prevent leakage of ions or water.

The few existing structures of Sec translocons have provided important insights into how this is all achieved. Van den Berg and colleagues determined a 3.2 Å resolution crystal structure of SecY from *Methanococcus jannaschii* [22]. This showed a heterotrimeric complex consisting of three subunits: α, β and γ (Figure 1.4). The α subunit is the largest, containing 10 TM helices. It seems to be composed of two halves, TM1-5 and TM6-10, with the γ subunit acting as a brace to keep the halves together. The area between TM2 and TM7 seems able to open laterally, with a hinge between TM5 and TM6; this area has been termed the 'lateral gate' (Figure 1.4). The central pore of the complex contains a ring of hydrophobic residues at its narrowest point, termed the 'pore ring'. The pore also seems blocked on the cytoplasmic side by a small helix termed the 'plug'.

These structural features immediately suggest how the translocon achieves some aspects of membrane protein insertion. It is reasonable to surmise that the central pore accommodates a polypeptide chain. The lateral gate provides a possible route for the chain to enter the membrane, while the pore ring and plug can help prevent ion and water leakage.

An X-ray structure of SecY from *Thermus thermophilus* [23] shows the complex in a 'pre-open' state where a hydrophobic crack opens to the cytoplasm, relative to the van den Berg structure. This crack is interpreted as the beginning of a conformational change triggered by the binding of a Fab fragment, mimicking the effect of SecA binding. (SecA transports



(a) Side view, lateral gate

(b) Top view, lateral gate

(c) Side view, pore ring

(d) Top view, pore ring

Figure 1.4: The crystal structure of SecY from *Methanococcus jannaschii* [22], PDB code 1RH5. The α,β and γ subunits are grey, cyan and purple respectively, and the plug is yellow. (a) and (b) the helices comprising the lateral gate are orange and red. (c) and (d) the atoms comprising the pore ring are shown in red.

an unfolded protein to the translocon complex during post-translational translocation.) A 4.5 Å structure of SecY bound to SecA [24] has also been produced, showing a 'two helix finger' of SecA inside the cytoplasmic funnel of SecY, interpreted as part of the mechanism to move the polypeptide chain through the translocon pore. A more recent 3.1 Å resolution structure of SecY from *Pyrococcus furiosus* shows the translocon in an open state, showing the plug still occluding the channel when the lateral gate is open [25].

This structural information, combined with data from other experiments, can be used to postulate a model for membrane protein insertion. Firstly, a length of polypeptide chain is synthesised in the ribosome. For co-translational translocation, the ribosome nascent chain (RNC) is recognised by a signal recognition particle (SRP), which guides the ribosome to bind with the translocon complex [26] (Figure 1.5). Inner membrane proteins in bacteria mainly use co-translational translocation, while soluble proteins typically employ post-translational translocation [27]. Chain elongation requires GTP hydrolysis, though polypeptide movement through the channel is independent of such hydrolysis [28]. Post-translational translocation is energized by hydrolysis of ATP and the protonmotive force [17], combined with either power-stroke [29] or Brownian ratchet [30] models for movement of peptide chain through the translocon. It is uncertain whether the polypeptide chain is pushed through the translocon by continued peptide synthesis in the ribosome for co-translational translocation [31, 32].

The mechanism by which the translocon selects which sections of the polypeptide chain to transfer into the membrane, and which into the cell, is not fully understood [34], although experiments from Gunnar von Heijne and colleagues have deciphered the molecular 'code' used. In these experiments the Leader Peptidase protein is used to challenge the Sec61 translocon with 19-residue H segments flanked by glycosylation sites (see Figure 1.6). The ratio of singly to doubly glycosylated H segments is interpreted as the probability distribution of inserted/not inserted helices, from which an apparent free energy of insertion (ΔG_{app}) can be derived. Early experiments calculated the contribution of each residue to ΔG_{app} [35], and so derived a hydrophobicity scale. Further experiments have revealed the positional-dependence of these residue insertion free energies [36].

The 'biological' hydrophobicity scale derived from these translocon-mediated experiments is surprisingly well-correlated with the Wimley-White water-octanol partitioning scale [37], derived from the partitioning of whole-residues between water and octanol. This has led to the suggestion that the partitioning of the polypeptide chain from the translocon

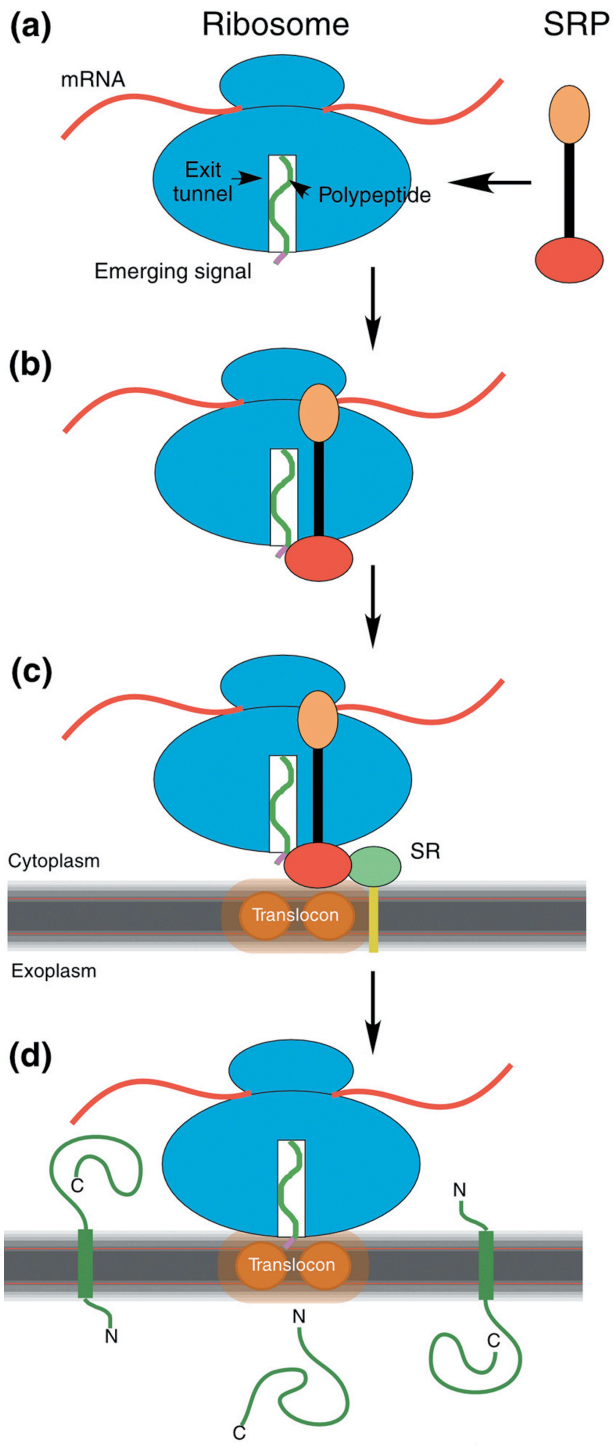


Figure 1.5: Insertion of TM helices by the ribosome-translocon complex. (a) The signal recognition particle (SRP) binds to the ribosome nascent chain (RNC), suspending further chain elongation (b). The ribosome/SRP complex then binds to the translocon (c), and the peptide chain is guided into the translocon pore (d). Figure adapted from [33].

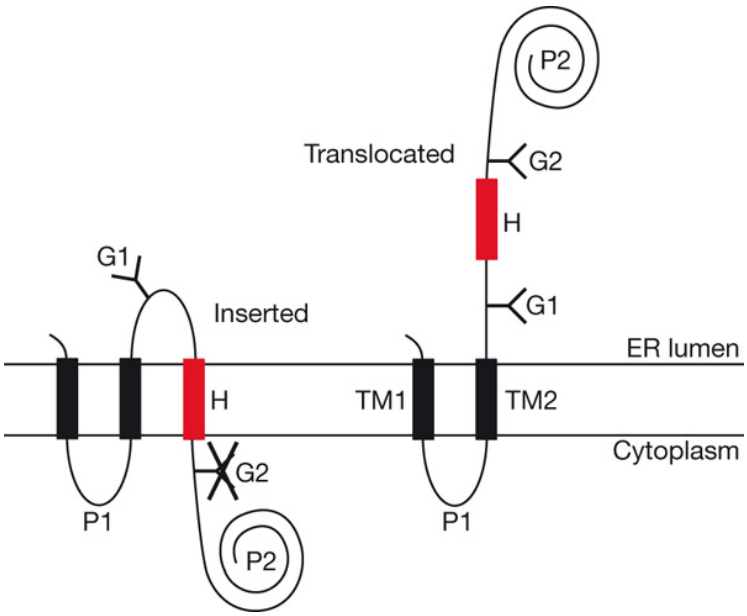


Figure 1.6: The system setup for translocon-mediated insertion experiments. The Leader Peptidase protein has two helices, TM1 and TM2, and a luminal domain P2. H-segments (red) are engineered into the P2 domain, flanked by two glycosylation sites G1 and G2. H segments with only G1 glycosylated are interpreted as adopting an inserted orientation; those with both G1 and G2 glycosylated are interpreted as adopting a non-inserted orientation. Figure taken from [36].

occurs by an equilibrium process, where the lateral gate of the translocon opens to allow the peptide to sample the hydrophobic membrane interior and choose whether to partition there or remain in the aqueous environment of the translocon pore. Polypeptide chains elongate on a timescale of milliseconds to seconds (faster in prokaryotes than eukaryotes) [38, 39, 40, 41]. This is much slower than the timescale over which the lateral gate of the translocon would be expected to open due to thermal fluctuations, so there is sufficient time for such a partitioning event. An alternative model for the selection of TM peptides in the translocon contends that the hydrophobicity of the peptide determines whether the translocon is stabilised in a state with the lateral gate either open or closed [42]. This would also be consistent with the correlation between the translocon-mediated and water/octanol partitioning scales.

1.5.2 Stage Two: Helix-Helix Aggregation

Once TM helices have entered the membrane, hydrophobic forces are unlikely to contribute significantly to their aggregation, as they are already in a hydrophobic environment. Instead, a variety of other forces stabilise their interactions. Van der Waals forces stabilise close packed helix-helix associations, often through the i, $i+4$ separation of small residues - Gly, Ala or Ser. Helix dimers known to interact via a GX_3G motif include Glycophorin

A (GpA) [43], major coat protein of Ff bacteriophage [44] and a variety of syndecans [45]. Other dimers interact via heptad motifs, including phospholamban [46] and the anthrax toxin receptor 1 [47]. Interactions between individual polar residues in the TM region can also result in dimerisation [48, 49, 50], while other sequences are stabilised by interhelical hydrogen bonds [51, 52, 53, 54]. Cation-π and/or π-π interactions between aromatic or basic residues are also observed in some TM dimers [55, 56]. For a recent review of helix-helix aggregation in the context of membrane protein folding, see [16] or [18].

1.6 Simulations of Membrane Proteins

Experimental methods such as X-ray crystallography can provide atomic resolution structural details of membrane proteins. However, these are typically static structures; experimental techniques investigating the dynamics of proteins, such as fluorescence resonance energy transfer (FRET), often involve adding artificial elements to the system which may influence the measurements made. Computer simulations provide a way of investigating the dynamic behaviour of these proteins, transplanted into lipid bilayers or micelles, without adding artificial elements. Computer simulations can therefore be complementary to experiments, both by explaining experimental observations in molecular detail and by making predictions which can be tested by experiment.

Computer simulations can be performed at different levels of granularity, which can provide insights into different types of behaviour (Figure 1.7). Hybrid quantum mechanics / molecular mechanics (QM/MM) methods are often employed to investigate systems where electrons are significant, such as enzymatic reactions [57, 58]. Examples of such systems where QM/MM modelling has proved useful include myoglobin [59], citrate synthase [60] and phenol hydroxylase [61].

Simulations at the atomistic level have provided useful insights for a variety of different membrane proteins. Mechanisms used by a variety of transporters have been illuminated, including the permeation of ions through channels [62, 63]; water through aquaporins [64]; the alternating access model for LacY [65, 66]; and conformational changes in the mitochondrial ADP/ATP carrier [67, 68] and LeuT [69]. Techniques such as dynamics importance sampling (DIMS) have been useful in mapping low-energy pathways between different states [70]; similarly metadynamics has been used to investigate processes such as ion permeation across the membrane and peptide folding [71]. The mechanisms underlying the function of potassium-gated (Kv) channels have been investigated by simulating the effect

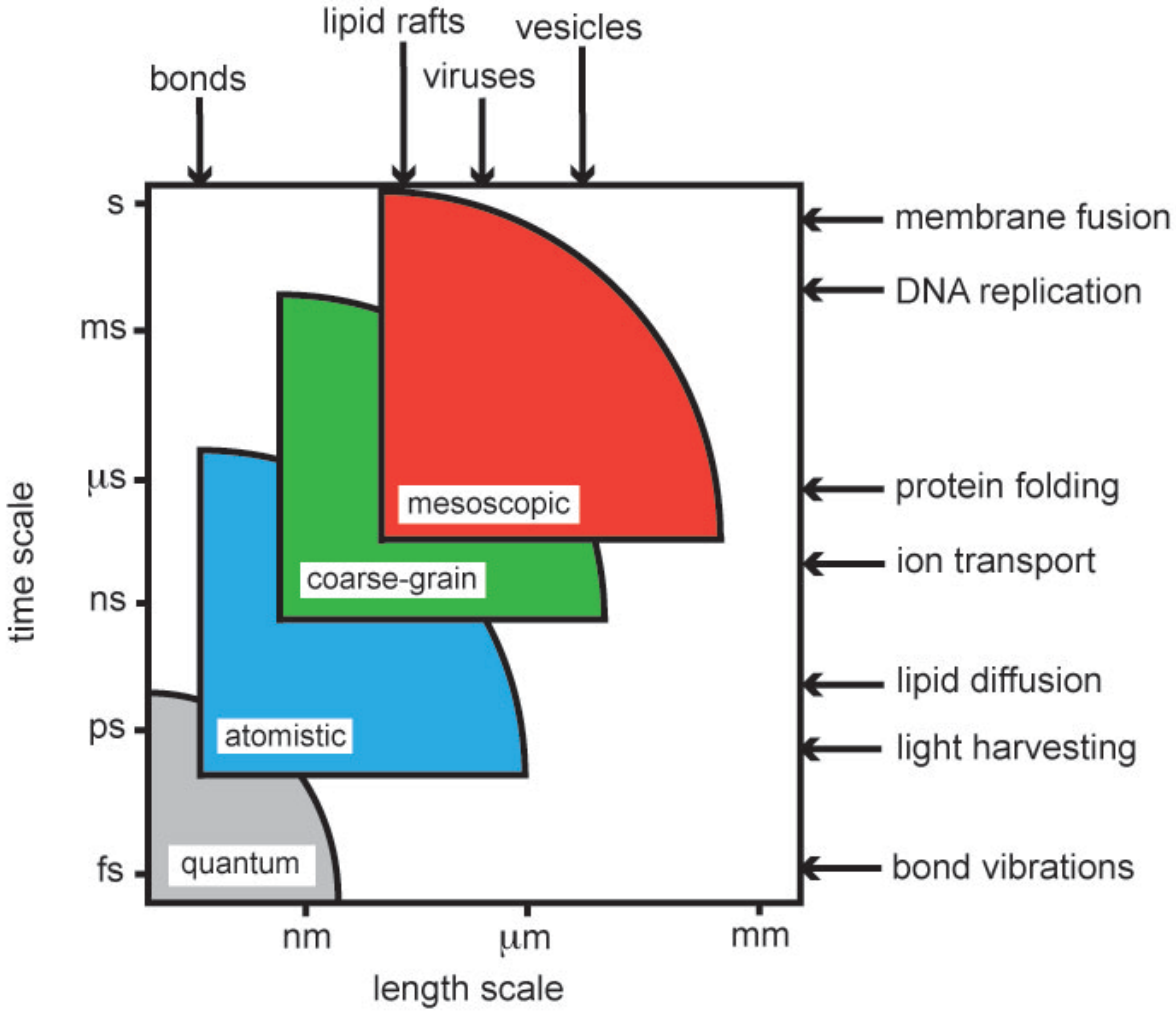


Figure 1.7: Approximate time and length scales accessible from different simulation methods. Some characteristic membrane events are also indicated.

of an electric field on the charged residues of the gating TM helices [72, 13, 73, 74]. A one-microsecond simulation of an open structure of a ligand-gated ion channel showed conformational changes which seem to result in closing of the channel [75]. Similar microsecond-scale atomistic simulations of a GPCR showed evidence of the formation of a functionally-important salt bridge, helping to resolve some controversy over the absence of this 'ionic lock' in static structures of the GPCRs [76, 77].

1.6.1 Coarse-Grained Simulations

QM/MM and atomistic simulations are limited to timescales of $\sim$1 microsecond, and system sizes of tens of nanometres. Many physiologically relevant processes occur at longer time and larger length scales, such as membrane fusion and DNA replication. Simulating these kinds of processes requires models with fewer details, so-called coarse-grained (CG) models. These usually either involve some kind of mean-field approximation or treating groups of atoms using a single interaction site, with internal degrees of freedom expressed as parameters of the model.

Mean-field theories consider the interaction of one particle via an average interaction derived from the effect of all other particles in the system, rather than attempting to calculate each of these interactions individually. Examples include chain-packing theory, which has been applied to the perturbation of membranes by TM proteins [78]; membrane elasticity theory, which has been applied to calculate the energy penalty of a mismatch between the hydrophobic thickness of a bilayer and of the TM region of an integral protein [79]; and Poisson-Boltzmann theory, which has been used to describe the energy of electrostatic interactions between charged proteins and charged lipids in a bilayer [80].

Another mesoscopic model for representing proteins comes from elastic network models (ENM). These represent proteins by a network of beads connected by elastic springs, ignoring chemical interactions [81, 82, 83, 84]. despite their simplicity ENM models have successfully reproduced the shape of the low-frequency part of the the density of states of the muscle protein G-actin, as well as fluctuations of its Cα atoms [85]. ENM models have also been useful in identifying functionally important modes for the chaperonin GroEL [86], and have predicted conformational changes triggered by ligand-binding for HCV NS3 helicase [87].

A wide variety of CG models use one particle to represent groups of atoms. Marrink et al. produced a semi-quantitative CG model for lipids where a 4-to-1 mapping of non-

hydrogen atoms is used [88]. This has been independently extended to apply to proteins by Bond et al. [1, 89, 90] and to lipoprotein particles by Shih et al. [91, 92]. The Bond model represents each amino acid by one backbone particle with between zero and two side chain particles (as this model was used extensively in this thesis it is discussed in more detail in Chapter 2). Marrink et al. recently extended their CG lipid model to apply to proteins [93, 94]; the updated model is referred to as MARTINI and has been widely applied to a variety of membrane proteins [95, 96, 97, 98, 99].

Other examples of CG models which have been developed to simulate biomolecules include a CG model for amino acids, developed by Devane et al. [100]; Markvoort et al. developed a similar CG model to study vesicles [101, 102, 103], and recently have developed a CG model for membrane proteins which includes a hydrogen bonding potential [104]. Orsi et al. produced a CG model for biomembranes which treats electrostatics explicitly, and uses the anisotropic Gay-Berne potential to model hydrocarbon tails [105, 106]. This model has also been used for hybrid simulations, where parts of a system are described at an atomistic level, while the rest is treated with a CG model [107]. Another hybrid all atom-CG system has been developed by Izvekov et al. [108, 109, 110, 111, 112, 113] and applied to a variety of systems [114, 115, 116, 117]. Berendt Smit and colleagues have used a CG model with a very low level of granularity, combined with dissipative particle dynamics (DPD), to investigate how lipids affect protein-protein interactions [118, 119, 120, 121].

1.7 Simulations of Helices

There are a variety of motivations for studying the interactions of individual helices, rather than more complex proteins, in a bilayer. Some individual helices perform biologically important functions, such as antimicrobial peptides [122, 123], peptidic venoms [124], and the TM helices of receptor tyrosine kinases [125]. The function of these helices can often be elucidated by direct study of their interactions with and within a bilayer. Other helices can be derived from dissecting an α-helical protein into its constituent TM domains; as the two stage model describes, these helices are often independently stable in the membrane, and are generally easier to produce and handle than membrane proteins [16]. Investigating these helices can often produce insights into the folding and function of the membrane proteins from which they were derived. Model peptides with artificial sequences can provide a controlled way of investigating factors contributing to TM helix stability, e.g. peptide length or hydrophobicity [126, 127].

A variety of TM helices have therefore been investigated using computational techniques. Implicit solvent techniques such as generalized Born electrostatics models [128] have been used to investigate TM helix insertion, providing a computationally efficient way of simulating a helix in atomistic detail [129]. The high thermostability of some helices has allowed simulations run at high temperatures to improve the sampling of helix insertion in fully atomistic detail, an approach which has confirmed the ability of simulations to closely match results from translocon-mediated insertion experiments [130, 131, 132]. The ability of computational simulations to provide a continuous 'movie' of a system has proved useful for investigating how helices and bilayers adapt to a mismatch between the hydrophobic thickness of the bilayer and that of the helix [133, 134].

The CG model in this thesis has been used extensively to investigate TM helices, and has been validated against a range of experimental data. CG simulations of the TM region of glycophorin A (GpA) match a structure derived from NMR spectroscopy, while simulations of a range of disruptive mutants have provided estimates of $\Delta\Delta G_{dimerisation}$ consistent with experiment [135]. Similar CG simulations of dimerisation of the TM region of both the syndecan-2 receptor and the αIIb/β3 integrin are consistent with experimentally derived structures [136]. High throughput CG methods have recently been developed, and have shown good correlation with translocon-mediated insertion experiments [137] and NMR experiments on the artificial GWALP peptides [138]. The permeation mechanism of the antimicrobial peptide maculatin has also been illuminated by CG simulations [139]. It is therefore apparent that the CG model used in this thesis can model TM helices in bilayers effectively, providing results both consistent with and complementary to experimental data; however, the simplifications in the model mean that care is required in selecting the appropriate system to model, and tractable problems to address. As with all computational methods the results should be validated against experimental data where possible.

1.8 Details of Work Presented in this Thesis

In Chapter 3 I present an investigation of the energetics of transmembrane helix insertion as calculated by CGMD simulations. This is compared with similar data from translocon-mediated insertion experiments, which raises questions about the interpretation of such experiments. The difficulty of interpreting the energetics of such a complex biological process as TM helix insertion via the translocon is discussed. Chapters 4 and 5 consider a method which may shed some light on the association of TM helices in a bilayer, a process related to

the second stage of membrane protein insertion. The systems studied are the TM regions of the receptor tyrosine kinases EphA1 and EphA2, chosen because there are NMR structures which suggest multiple interaction modes for both systems [140, 141]. Chapter 4 describes unrestrained CGMD simulations of both systems, with atomistic simulation equivalents. Chapter 5 describes calculations of the energetics of the dimerisation process for both systems using CGMD simulations. Although the results are shown to be comparable with a variety of similar calculations performed in other groups, the various sources of error in the calculations suggest that a confident estimate of the energy difference between multiple interaction states is not currently possible.

Chapter 2

Theory & Methods

2.1 Introduction

This chapter contains an overview of the theory behind the computational methods used in this thesis. Computer simulations can provide a link between the macroscopic world of experiment and the microscopic interactions of the molecules within systems of interest. In particular, molecular dynamics (MD) has been extensively used to further understand the relationship between structure and function of membrane proteins [142, 143, 144, 145]. In this chapter I discuss the theory behind MD, focusing particularly on the coarse-grained model used for many simulations in this thesis. I also discuss the theory behind free energy calculations using computational methods, with particular reference to the umbrella sampling method for enhanced sampling.

2.2 Molecular Dynamics Simulations

2.2.1 The Born-Oppenheimer Approximation

Simulation methods involving quantum mechanics (QM) calculations can provide extremely accurate descriptions of the dynamics of molecular systems; however, they are computationally very expensive. The Born-Oppenheimer approximation [146] holds that electronic motion can be decoupled from motion of nuclei. The significantly smaller mass of electrons means that they can adjust to changes in the positions of nuclei over very short timescales with respect to the nuclei motion. It is therefore reasonable to consider the energy of a molecule in its ground electronic state as a function only of the position of the nuclei. Atomistic molecular dynamics simulations usually do not involve calculating the motion of elec-

trons, so implicitly rely on the Born-Oppenheimer assumption.

2.2.2 Solving Newton's Equations

In molecular mechanics (MM) atoms are treated as van der Waals spheres with a point charge at the centre. All atoms in the system act under a potential $\mathcal{V}(\mathbf{r})$. The force $\mathbf{F}$ on a particle is related to the potential it experiences through

$$\mathbf{F} = -\frac{d\mathcal{V}(\mathbf{r})}{d\mathbf{r}} \tag{2.1}$$

with $\mathcal{V}(\mathbf{r})$ the potential on the particle at position $\mathbf{r}$. This can be used in conjunction with Newton's second law

$$\mathbf{F} = m\frac{d^2\mathbf{r}}{dt^2} \tag{2.2}$$

to calculate the position of the particle as a function of time. Thus if we know the potential a particle experiences we can (with boundary conditions), in principle, know its position also, to a classical approximation.

Newton's second law can be used to derive a set of equations relating the acceleration $\mathbf{a}$, velocity $\mathbf{v}$ and position $\mathbf{r}$ of a particle:

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \frac{\mathbf{F}}{m}, \tag{2.3}$$

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \mathbf{a}t + \mathbf{v}_0, \tag{2.4}$$

$$\mathbf{r} = \mathbf{v}_0 t + \frac{1}{2}\mathbf{a}t^2 + \mathbf{r}_0. \tag{2.5}$$

From these equations it is apparent that knowledge of initial positions $(\mathbf{r}_0)$ and initial velocities $(\mathbf{v}_0)$ is sufficient to calculate the subsequent positions of all atoms as a function of time. Initial positions are usually obtained from X-ray / NMR structures of proteins, or constructed models. Initial velocities are taken from a temperature-dependent Maxwell-Boltzmann distribution of velocities:

$$p(\mathbf{v}) = \sqrt{\frac{m}{2\pi k_B T}} e^{\left(-\frac{m\mathbf{v}^2}{2k_B T}\right)} \tag{2.6}$$

In general equation 2.2 cannot be solved analytically, so is typically solved for MD simulations using finite difference methods. If the force $\mathbf{F}(t)$, position $\mathbf{r}(t)$ and velocity $\mathbf{v}(t)$ are known for a given time t, finite difference methods calculate these same values for a new time $t+\triangle t$. The finite difference method used in all simulations in this thesis is the leap-frog

algorithm [147]. This can be derived from the Taylor series expansions

$$\mathbf{r}(t+\triangle t) = \mathbf{r}(t+\frac{\triangle t}{2}) + \frac{d\mathbf{r}(t+\frac{\triangle t}{2})}{dt}\frac{\triangle t}{2} + \frac{d^2\mathbf{r}(t+\frac{\triangle t}{2})}{dt^2}\frac{\triangle t^2}{4} + \dots \tag{2.7}$$

and

$$\mathbf{r}(t) = \mathbf{r}(t+\frac{\triangle t}{2}) - \frac{d\mathbf{r}(t+\frac{\triangle t}{2})}{dt}\frac{\triangle t}{2} + \frac{d^2\mathbf{r}(t+\frac{\triangle t}{2})}{dt^2}\frac{\triangle t^2}{4} - \dots \tag{2.8}$$

Taking equation 2.8 from equation 2.7 and neglecting terms $\mathcal{O}(\triangle t^4)$ we arrive at

$$\mathbf{r}(t+\triangle t) = \mathbf{r}(t) + \frac{d\mathbf{r}(t+\frac{\triangle t}{2})}{dt}\triangle t \tag{2.9}$$

A similar pair of Taylor expansions of $\mathbf{v}(t+\frac{\triangle t}{2})$ and $\mathbf{v}(t-\frac{\triangle t}{2})$, neglecting terms $\mathcal{O}(\triangle t^2)$, leads to

$$\frac{d\mathbf{r}(t+\frac{\triangle t}{2})}{dt} = \frac{d\mathbf{r}(t-\frac{\triangle t}{2})}{dt} + \frac{d^2\mathbf{r}(t+\frac{\triangle t}{2})}{dt^2}\triangle t \tag{2.10}$$

Equations 2.9 and 2.10 can be used to iteratively calculate $\mathbf{r}(t)$ and $\mathbf{v}(t+\frac{\triangle t}{2})$. The leapfrog algorithm has the advantage over many similar methods that it explicitly calculates the velocity and does not require the calculation of differences of large numbers. Other finite difference methods are available, including the velocity Verlet algorithm [148] and Beeman's algorithm [149].

In MD simulations $\triangle t$ is the timestep of the simulation, which needs to be smaller than the shortest frequency vibrations in the system. In atomistic simulations the shortest frequency vibrations are typically bond vibrations (particularly bonds involving hydrogen atoms), limiting the timestep to the order of one femtosecond. Coarse grained simulations which limit bond vibrations can use a much longer timestep.

2.2.3 Force fields

Section 2.2.2 showed that we can, in principle, know the time evolution of the position of a system of particles if we know the potential which acts on them. In MM force fields provide the functional form of this potential, together with a set of parameters for different atom types. The potential function is formed by the sum of bonded and non-bonded interactions:

$$\mathcal{V}(\mathbf{r}) = \mathcal{V}_{bonded} + \mathcal{V}_{nonbonded}. \tag{2.11}$$

Bonded Interactions

The bonded interactions are comprised of the sum of covalent bond stretching ($\mathcal{V}_{bonds}$), bond angle bending ($\mathcal{V}_{angles}$), proper dihedral torsions ($\mathcal{V}_{dihedral}$) and improper torsions ($\mathcal{V}_{improper}$).

Typically the potential energy contributions from bond lengths and angles are treated as harmonic potentials:

$$\mathcal{V}_{bonds} = \sum_{bonds} \frac{k_b}{2}(b - b_0)^2 \tag{2.12}$$

$$\mathcal{V}_{angles} = \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{2.13}$$

where the force constants k_b and k_θ determine the energy cost of deviation from the reference bond lengths and angles (b_0 and θ_0 respectively). These harmonic potentials are a good approximation only over a limited range of lengths and angles. Other functional forms are possible, for example the Morse potential [150], but harmonic potentials are computationally efficient and valid over the range of bond lengths and angles typically experienced during MD simulations of membrane proteins.

The dihedral torsion potential represents the change in energy as a bond rotates. This is typically represented by a cosine sum, for example

$$\mathcal{V}_{dihedral} = \sum_{dihedral} \frac{k_\chi}{2}(1 + cos(n\chi - \gamma)) \tag{2.14}$$

where k_χ is a force constant, n is the multiplicity (number of minima in $180°$), χ is the torsion angle and γ is a phase factor.

An 'improper' dihedral potential is also included to treat out-of-plane bending motions (e.g. to maintain planarity in a molecular structure). This also typically takes a harmonic form:

$$\mathcal{V}_{improper} = \sum_{impropers} \frac{k_\varphi}{2}(\varphi - \varphi_0)^2 \tag{2.15}$$

with k_φ a force constant and φ_0 a reference angle.

Non-bonded Interactions

Non-bonded interactions consist of the sum of Van der Waals (VDW) and electrostatics interactions.

$$\mathcal{V}_{nonbonded} = \mathcal{V}_{VDW} + \mathcal{V}_{Coulomb} \tag{2.16}$$

These non-bonded interactions only apply to pairs of atoms that are separated by at least three covalent bonds, or are part of different molecules.

The standard Lennard-Jones potential is typically used to model VDW interactions and steric repulsion. This takes the form

$$\mathcal{V}_{VDW} = \sum_{i \neq j} 4\epsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] \tag{2.17}$$

where r_{ij} is the distance between the centres of mass of the two atoms i and j, σ_{ij} is the distance at which the VDW interaction is zero, and ϵ_{ij} is the well depth. The two elements to the potential are the attractive forces at long range arising from dispersive forces, and a shorter range steric repulsion. Alternative forms for VDW interactions are available, such as the Buckingham potential [151, page 209], but are generally computationally more expensive.

Coulombic interactions are modelled with the standard potential

$$\mathcal{V}_{coulomb} = \sum_{i \neq j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.18}$$

with q_i and q_j the charges on the two atoms, r_{ij} their centre of mass separation and ϵ_0 the relative permittivity of free space.

Parameterisation

The first stage in parameterising a force field is choosing a data set to fit the parameters to. Ideally this would be experimental data, but when this is unavailable the results from quantum mechanics (QM) calculations are often used. Once the data set has been chosen, there are two approaches to fitting the parameters. One is a simple trial and error, where the dependence of each parameter on others means that the iteration has to be done in stages. The second approach is to use the least-squares method to minimise the difference between calculated and experimental data.

2.2.4 Periodic Boundary Conditions

Computational resources dictate the number of particles in a system which can be realistically simulated. The edges of this simulation present a difficulty: a 'wall' would produce

non-realistic forces, whereas a vacuum would lead to parts of the simulation evaporating. In all simulations in this thesis periodic boundary conditions are therefore applied. The system is treated as though it is surrounded by repeating identical boxes in all directions; thus if a molecule were to diffuse out of the left hand side of the box, its image molecule would reappear on the right. This ensures that the number of particles in the system remains constant while allowing the solvent particles to behave as though in bulk solution.

2.2.5 Energy Minimization

Initial configurations for MD simulations often contain unfavourable steric clashes, which may cause instabilities in a simulation. It is therefore important to energy minimise initial configurations, particularly if new molecules have been added. Formally, this corresponds to finding the coordinates $\mathbf{r}_{min}$ such that the potential energy function $\mathcal{V}(\mathbf{r})$ satisfies

$$\frac{d\mathcal{V}(\mathbf{r}_{min})}{d\mathbf{r}} = 0; \frac{d^2\mathcal{V}(\mathbf{r}_{min})}{d\mathbf{r}^2} > 0. \tag{2.19}$$

The complexity of the energy landscape of biomolecular MD systems means that an analytical solution is usually not possible, so numerical methods are used instead. Simulations in this thesis used the steepest descent method [151, page 262] or limited memory Broyden-Fletcher-Goldfarb-Shanno quasi Newton minimiser (l-bfgs) [152, page 46]. The steepest descent method moves particles in the direction of the force vector defined through

$$\mathbf{F} = -\frac{d\mathcal{V}(\mathbf{r})}{d\mathbf{r}} \tag{2.20}$$

by a certain step size λ. If the iteration produces a reduction in energy, then λ is increased. If the energy is increased by the iterative step, then λ is decreased. Each iterative step is orthogonal to the previous one. This is generally robust but is inefficient on an irregular energy landscape. The original bfgs method creates successively better approximations of the inverse Hessian matrix[1], moving the system to its currently estimated minimum. The l-bfgs method reduces the amount of memory needed by approximating the inverse Hessian matrix by a fixed number of corrections from previous steps. This is more efficient when close to the local minimum than the steepest descents method.

[1]The Hessian matrix is the matrix of second derivatives; it describes the local curvature of the energy function.

2.2.6 Temperature and Pressure Coupling

The simulations described in this thesis were all performed in the NPT ensemble: keeping constant the number of particles (N), pressure (P), and temperature (T). These were achieved by coupling to a Berendsen thermostat and barostat. The Berendsen thermostat [153] couples the temperature of the system $T(t)$ to an external heat bath, at a reference temperature T_{bath}, which prevents too much heat from entering or leaving the system. Different parts of the system - protein, lipid and solvent - are coupled separately to prevent temperature imbalances accruing in the simulation. The rate of change of the temperature is proportional to the temperature difference between the simulation and the reference temperature:

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_{bath} - T(t)) \tag{2.21}$$

where τ_T is the temperature coupling constant, which determines how tightly the simulation is coupled to the bath.

Berendsen pressure coupling works in an analogous way, with the pressure of the system $P(t)$ coupled to an external pressure 'bath' at pressure P_{bath}. The rate of change of pressure is proportional to the difference between the pressure in the system and that of the bath:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p}(P_{bath} - P(t)) \tag{2.22}$$

where τ_p is the pressure coupling constant. Box vectors are coupled semi-isotropically, where the vectors perpendicular to the bilayer normal are rescaled together, with the vector parallel to the bilayer normal varying separately. This helps to maintain a constant surface area per lipid ratio. There are other available thermostats (for example Nose-Hoover [154, 155]) and barostats (for example Parrinello-Rahman [156, 157]).

2.2.7 Efficient Electrostatics

Coulomb interactions decay with distance much less rapidly than VDW interactions. It is possible to introduce a larger cutoff for Coulomb interactions, but this can introduce instabilities into the simulation. Atomistic simulations in this thesis instead used the Particle Mesh Ewald (PME) method [158, 159]. The principle is to surround the point charges in the system with Gaussian charge distributions of equal but opposite charge. These distributions screen the effects of the point charges so the net effects decay rapidly in real space, and a cutoff can be introduced with limited loss of accuracy. Further charge distributions

can then be added to cancel these distributions, and the net effect of these can be summed in reciprocal space, where rapid decay again allows for the introduction of a cutoff. These sums in reciprocal space can be calculated efficiently using the Fast Fourier Transform (FFT) method. An alternative to PME is the reaction field method, where the electrostatic interactions of any atom with all atoms within a cutoff distance are calculated explicitly. Beyond the cutoff electrostatic interactions are treated through a homogeneous medium of a certain dielectric constant [160].

2.2.8 Bond Constraints

The motions of interest in a biological system typically come from changes in torsion angles and non-bonded interactions, not from bond stretching. Bond vibrations normally occur at a very high frequency; constraining[2] bond lengths to their reference length allows longer timesteps and is thus computationally efficient. The algorithm LINCS [161] was used in atomistic simulations in this thesis to update the reference bond length at every timestep.

2.2.9 Position Restraints

Restraining a part of a system while another part equilibrates is a sometimes necessary step for generating a stable simulation. These restraints are typically harmonic, with a particle at position **r** restrained to a reference position $\mathbf{R}_{ref}$ by the potential

$$V_{pr} = k_{pr}(\mathbf{r} - \mathbf{R}_{ref})^2. \tag{2.23}$$

2.3 Coarse-Grained Simulations

2.3.1 Coarse-Graining

The calculation of non-bonded interactions in MD simulations is computationally expensive. Reducing the number of calculations necessary can be achieved by using coarse-grained models, which treat groups of (typically bonded) atoms as a single interaction site. Appropriate choices of parameters for CG models can ensure that the essential chemistry of a system is retained, while increasing computer speed by orders of magnitude. This allows access to greater time and length scales of simulation compared with models which treat all atoms explicitly.

[2]Systems are required to satisfy constraints; restraints allow some deviation.

Some 'united atom' force fields treat hydrogen atoms as part of a single particle with adjacent bonded atoms. For example, a methyl group may be represented as a single interaction site, with the mass of the group and non-bonded interactions taking into account the implicit hydrogens [162].

Models which reduce the level of detail in the system even further have been extensively used to study the relationship between structure and function for membrane proteins [163, 120].

2.3.2 The CG Lipid and Protein Model

The CG model used in this thesis was developed by Bond et al. [1, 90, 89], and was based on a model originally developed for lipid and surfactant systems by Marrink et al. [88]. This model maps roughly four non-hydrogen atoms on to a single interaction site. There are four different particle types: polar (P); mixed polar/apolar (N); hydrophobic apolar (C); and charged (Q). A comparison of a DPPC molecule in an atomistic representation and in a CG representation is shown in Figure 2.3.2. Additional subtypes are defined for N and Q particles to reflect hydrogen bonding capabilities (donor, acceptor, both or neither). There are five levels of Lennard-Jones interaction, ranging from attractive through intermediate to (effectively) repulsive. Reducing the number of particles, using cut-offs for long range interactions and longer timesteps allow this model to achieve a speedup of between three and four orders of magnitude compared with fully atomistic simulations. The lipid model was parameterised with respect to densities of liquid alkanes as well as structural and dynamic bilayer properties such as area per headgroup, width, bending modulus and lipid lateral diffusion coefficient [88].

Bond et al. adapted this model to apply to membrane proteins [90, 89, 1]. One CG particle is used to represent a peptide bond, and zero, one or two particles represent an amino acid side chain. An additional particle type Nm was introduced to more accurately reflect the polarity of a helix backbone. The particle types for amino acid sidechains are shown in Table 2.1. The Lennard-Jones particle size is 0.47 nm. The force constant for protein bonds is 5000 kJ mol^{-1} nm^{-2}, with a reference bond length between backbone particles of 0.36 nm. Reference bond lengths and angles for amino acid particles were chosen on the basis of comparison to a selection of high-resolution membrane protein crystal structures [1]. Thus for example although glutamate and aspartate both have the same side chain particle types (Table 2.1) their reference bond lengths were 0.36 nm and 0.26 nm respectively. Hydrogen

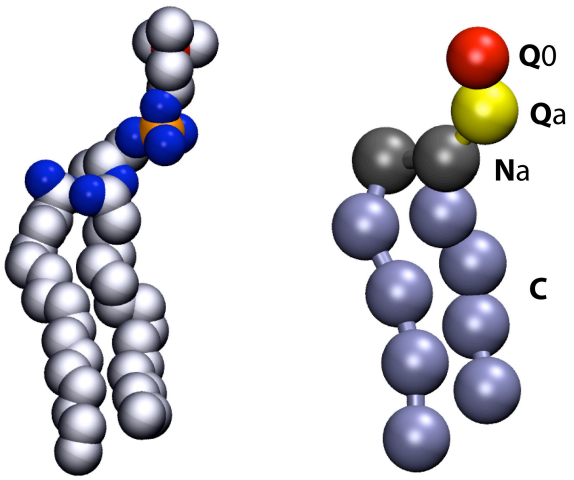


Figure 2.1: A comparison of a DPPC molecule shown in an atomistic representation (left) and in a CG representation (right). In the atomistic representation the carbon, phosphorous, nitrogen and oxygen atoms are white, red, orange and dark blue respectively. In the CG representation the choline, phosphate, glycerol and carbonyl moieties are red, orange, grey and iceblue respectively.

bonds in α-helices are mimicked using a force constant of 1000 kJ mol^{-1} nm^{-2} and reference bond length of 0.6 nm. The LJ interaction potential is smoothly shifted to zero between 9 Å and 12 Å. The Coulomb potential is smoothly shifted to zero between 0 Å and 12 Å.

The protein parameters were developed through an iterative procedure comparing free energies of insertion from thermodynamic integration simulations between (CG) water and (CG) cyclohexane, and the equivalent experimental values as obtained by Radzicka and Wolfenden [164]. The resulting correlation coefficient was 0.93 [1]. Values for the parameters for e.g. reference bond lengths and angles for different secondary structure elements were in good agreement with those obtained for the similar MARTINI model [94, 93], which was parameterised by estimating the equilibrium densities of low concentrations of CG particles dissolved in a water/butane two-phase system. The correlation with experimental values was robust to the choice of reference bond length, with the variation being $<$0.5 k_BT when the reference bond length was between half and double the chosen value [1].

This model has been used to investigate membrane protein-lipid interactions. Earlier variations on the model reproduced the lipid-protein interactions observed by a variety of noncrystallographic experimental techniques (site-directed chemical modification, intermolecular cross-linking, site-directed spin labelling) applied to a variety of membrane proteins with differing structural features [165]. More quantitative assessment of the CG model

amino acid	particle type
Ala (A)	**C**
Arg (R)	**N***d*+**Q***d* (charge +1)
Asn (N)	**P**
Asp (D)	**Q***a* (charge -1)
Cys (C)	**N***d*
Glu (E)	**Q***a* (charge -1)
Gln (Q)	**P**
Gly (G)	n/a
His (H)	**N***d*+**N***da*
Ile (I)	**C**
Leu (L)	**C**
Lys (K)	**C**+**Q***d* (charge +1)
Met (M)	**N***d*
Phe (F)	**C**+**C**
Pro (P)	**C**
Ser (S)	**N***da*
Thr (T)	**N***da*
Trp (W)	**C**+**N***d*
Tyr (Y)	**C**+**P**
Val (V)	**C**

Table 2.1: Amino acid sidechain particle types for the CG model used in this thesis [1].

was provided by comparisons between PMF calculations translating the CG amino acid side chains into a lipid bilayer, and similar atomistic PMF calculations [1]. These correlated extremely well with the transfer free energies from thermodynamic integration (correlation coefficient 0.99), and well with the translocon-mediated hydrophobicity scale (0.83).

2.4 Free Energy Calculations

Free energies are quantities of great interest to biochemists because in a system at equilibrium one of the free energies will be at a minimum. In particular the Gibbs free energy (G) - the amount of work available in a system under constant NPT - is useful as most experiments (and all the simulations in this thesis) are at constant NPT. G is defined as

$$G = H - TS = U + PV - TS, \tag{2.24}$$

where H is enthalpy, U internal energy, S entropy and T temperature. G may also be interpreted as related to the probability that a system will occupy a given state.

G is at a minimum at equilibrium under conditions of constant NPT. Another interpretation of this is that the sign of the change in G during a reaction, $\triangle G$, indicates whether a reaction will occur spontaneously (i.e. without a further external input of energy).

In principle finding the free energy of a system from a simulation requires knowledge of the complete phase space accessible to the system. This is usually impossible, though finding the difference in free energy between two states is often a tractable problem, as it involves finding the relative probabilities of occupation of the two different states.

Unrestrained MD simulations are usually not adequate to estimate free energy, as they will not adequately sample high energy regions of phase space, and may be unable to overcome kinetic barriers within obtainable computational timescales. A variety of methods have been developed to overcome these sampling problems, including umbrella sampling [166, 167], free energy perturbation (FEP) [168], thermodynamic integration [169], metadynamics [170], adaptive biasing force (ABF) [171] and steered MD using Jarzynski's equality [172]. See section 2.4.4 for further details.

2.4.1 The Potential of Mean Force

Notation and definitions in this section are taken from reference [167]. The potential of mean force (PMF) $\mathcal{W}(\xi)$ along some chosen coordinate ξ can be defined as

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right], \tag{2.25}$$

with ξ^* an arbitrary constant on the chosen coordinate. $\langle \rho(\xi) \rangle$ is the average distribution function, defined as

$$\langle \rho(\xi) \rangle = \frac{\int \delta(\xi'(\mathbf{R}) - \xi) e^{-U(\mathbf{R})/k_B T} d\mathbf{R}}{\int e^{-U(\mathbf{R})/k_B T} d\mathbf{R}}, \tag{2.26}$$

where $U(\mathbf{R})$ is the total energy of the system as a function of the general coordinates $\mathbf{R}$. $\xi'(\mathbf{R})$ is a function in general depending on a few or several degrees of freedom in the dynamical system, e.g. an angle, distance, or a more complicated function of the coordinates of the system. In practice $\xi'(\mathbf{R})$ represents the state of the restrained elements of the system.

2.4.2 Umbrella Sampling

Umbrella sampling is a technique that was developed by Torrie and Valleau [166] as a way of ensuring that all areas of the energy landscape along a chosen reaction coordinate could be sampled adequately. The idea is to add a biasing potential to the simulation to ensure that a subset of atoms in the system remain near the centre of the biasing potential. In practice, this is usually applied as a series of potentials applied to independent simulations (called windows), with the subsequent results recombined to form the PMF. A variety of methods

exist to unbias the data - in this thesis the weighted histogram analysis method (WHAM) was used (see section 2.4.3 for more details).

One reasonable choice for the form of the biasing potential is a harmonic restraint. In the i^{th} window this takes the form:

$$w_i(\xi) = \frac{k_w}{2}(\xi - \xi_i)^2 \tag{2.27}$$

where ξ_i is the centre of the window potential. This was the form of restraint used in all restrained simulations in this thesis.

The choice of reaction coordinate in umbrella sampling calculations is extremely important. It is directed by structural and biophysical information about the system of interest, and corresponds to the degree of freedom(s) on to which the free energy landscape is projected. The reaction coordinate may or may not be a path which the system would physically follow but must be defined such that all degrees of freedom 'perpendicular' can be adequately sampled. It can be a simple function of Cartesian coordinates or a more complex function of generalised coordinates. The PMF profile generated must be understood in the context of the reaction coordinate and the more complex multi-dimensional free energy landscape of the system.

For umbrella sampling to be effective, there must be sufficient overlap between consecutive restrained simulations such that the reaction coordinate is sampled fully. This can be tested by viewing histograms of the position of the restrained atoms as a function of the reaction coordinate. It is also necessary for all degrees of freedom that are not restrained - i.e. those 'perpendicular' to the reaction coordinate - to equilibrate sufficiently. In practice it is difficult to be certain that this has happened without extending the windows indefinitely, so this can be a substantial source of systematic errors. Checking that different parts of the simulations recombine to form the same PMF, that the PMF does not change as the simulations are extended, and using chemical intuition as to the likely preferred state of the system, are all ways of assessing adequate sampling.

2.4.3 Weighted Histogram Analysis Method (WHAM)

This explanation largely follows that provided by Roux in reference [167]. Consider the i^{th} window of a biased ensemble, acting under a potential $U(\mathbf{R}) + \omega_i(\xi)$. From this simulation we can obtain

$$\langle \rho(\xi) \rangle_i^{biased} = e^{-\omega_i(\xi)/k_BT} \frac{\langle \rho(\xi) \rangle^{unbiased}}{\langle e^{-\omega_i(\xi)/k_BT} \rangle} \tag{2.28}$$

where

$$\langle \rho(\xi) \rangle^{unbiased} = \frac{\int \delta(\xi'[\mathbf{R}] - \xi) e^{-U(\mathbf{R})/k_B T} d\mathbf{R}}{\int e^{-U(\mathbf{R})/k_B T} d\mathbf{R}}. \tag{2.29}$$

The unbiased PMF from the i^{th} window is then

$$\mathcal{W}_i(\xi)^{unbiased} = \mathcal{W}(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle_i^{biased}}{\langle \rho(\xi^*) \rangle} \right] - \omega_i(\xi) + F_i \tag{2.30}$$

where the unknown constant F_i is defined from

$$e^{-F_i/k_B T} = \langle e^{-\omega_i(\xi)/k_B T} \rangle \tag{2.31}$$

and represents the additional free energy associated with introducing the window potential.

At this stage a traditional approach is to find the unknown constants F_i by adjusting the various $\mathcal{W}_i(\xi)$ of adjacent windows in the region they overlap until they match. However, this requires a significant overlap, which means that much data is effectively unused. WHAM aims to utilise all information obtained in the simulations, and has the advantage of being easily generalisable to a multi-dimensional PMF.

Consider an umbrella sampling calculation involving N_w biased window simulations. Then an optimal estimate for the unbiased probability distribution function is a ξ-dependent weighted sum over the N_w individual unbiased distribution functions $[\langle \rho(\xi) \rangle]_i^{unbiased}$,

$$\langle \rho(\xi) \rangle^{unbiased} = \sum_{i=1}^{N_w} [\langle \rho(\xi) \rangle]_i^{unbiased} \times \left[\frac{n_i e^{-[\omega_i(\xi) - F_i]/k_B T}}{\sum_{j=1}^{N_w} n_j e^{-[\omega_j(\xi) - F_i]/k_B T}} \right] \tag{2.32}$$

where n_i is the number of independent data points obtained from the biased window simulation. From equations 2.28 and 2.31, the individual unbiased distribution function is

$$[\langle \rho(\xi) \rangle]_i^{unbiased} = [\langle \rho(\xi) \rangle]_i^{biased} \times e^{\omega_i - F_i/k_B T} \tag{2.33}$$

and equation 2.32 can be rewritten as

$$\langle \rho(\xi) \rangle^{unbiased} = \sum_{i=1}^{N_w} n_i \langle \rho(\xi) \rangle_i^{biased} \times \left[\sum_{j=1}^{N_w} n_j e^{-[\omega_j(\xi) - F_j]/k_B T} \right]^{-1}. \tag{2.34}$$

The free energy constants F_i can be determined from equation 2.31 using the optimal esti-

mate for the distribution function

$$e^{-F_i/k_BT} = \int e^{-\omega_i(\xi)/k_BT} \langle \rho(\xi) \rangle^{unbiased} d\xi. \tag{2.35}$$

Equations 2.34 and 2.35 are then solved self-consistently for the distribution function $\langle \rho(\xi) \rangle^{unbiased}$ and the free energy constants F_i. In practice this usually involves an initial guess for the free energy constants, allowing an estimate of the distribution function from equation 2.34. This estimate for the distribution function is then fed into equation 2.35 to provide a new estimate for the free energy constants. This process is repeated until the change of estimate for the distribution function after one iteration is below a certain threshold.

2.4.4 Other Methods for Calculating Free Energies

Thermodynamic Integration

Consider a coupling parameter λ which defines a path between two states of a system, corresponding to $\lambda = 0$ and $\lambda = 1$. The free energy (A) change with respect to λ is then

$$\frac{\partial A}{\partial \lambda} = \iint \frac{1}{Q(\lambda)} \frac{\partial \mathcal{H}(\mathbf{x}, \mathbf{p}_x)}{\partial \lambda} \exp\left[-\mathcal{H}(\mathbf{x}, \mathbf{p}_x)/k_BT\right] d\mathbf{x}\, d\mathbf{p}_x \tag{2.36}$$

where $\mathcal{H}$ is the Hamiltonian of the system and $Q(\lambda)$ is the partition function, given by

$$Q(\lambda) = \iint \exp\left[-\frac{\mathcal{H}(\mathbf{x}, \mathbf{p}_x)}{k_BT}\right] d\mathbf{x}\, d\mathbf{p}_x. \tag{2.37}$$

The PMF can be calculated from equation 2.36 using

$$\frac{d\mathcal{W}(\xi)}{d\xi} = -\frac{\partial A}{\partial \lambda}. \tag{2.38}$$

The free energy difference between two states can be found by integrating equation 2.36 between $\lambda = 0$ and $\lambda = 1$:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \langle \frac{d\mathcal{H}(\lambda)}{d\lambda} \rangle_\lambda d\lambda. \tag{2.39}$$

For a full derivation of the thermodynamic integration equations see [151, page 630].

Free Energy Perturbation

To outline the Free Energy Perturbation (FEP) technique [173] we introduce a similar thermodynamic coupling parameter λ as seen in thermodynamic integration. λ describes in-

termediate states between two states, corresponding to $\lambda = 0$ and $\lambda = 1$. FEP provides an estimate for the free energy change $\Delta A = A_1 - A_0$ by the sum of successive steps with intermediate values of λ. The equation for the free energy change is

$$\Delta A = -k_B T \langle e^{-\Delta U / k_B T} \rangle_{\lambda=0}^{eq} \tag{2.40}$$

where the average is as in equation 2.39, evaluated at $\lambda = 0$ and ΔU is the potential energy change. In practice free energy changes measured by FEP must be small, so the the free energy difference between states A and B is approximated by

$$\Delta A \approx -k_B T \ln \frac{1}{N} \sum_{i=1}^{N} e^{-\Delta U(p_i, x_i)/k_B T} \tag{2.41}$$

where p_i and x_i are momenta and positions of particle i in a sampled microstate and N is the number of sampled microstates.

FEP can be calculated bidirectionally to check convergence, using

$$\Delta A = k_B T \langle e^{\Delta U / k_B T} \rangle_{\lambda=1}^{eq}. \tag{2.42}$$

Steered MD using Jarzynski's Equality

Jarzynski's equality [174] describes a relationship between an equilibrium free energy difference between two states and work done through nonequilibrium processes during the transition [172]. Consider a process which changes the external coupling parameter λ from λ_0 at time zero to λ_t at time t. Then Jarzynski's equality states that, regardless of the speed of transition,

$$e^{-\frac{\Delta F}{k_B T}} = \langle e^{-\frac{E}{k_B T}} \rangle \tag{2.43}$$

where ΔF is the Helmholtz free energy difference between the two states and E is the work done during the transition.

To apply this to the calculation of the potential of mean force $\mathcal{W}$, add an external parameter λ such that λ is correlated with the reaction coordinate ξ. This can be done with the spring potential

$$h(\mathbf{r}; \lambda) = \frac{k}{2} [\xi(\mathbf{r}) - \lambda]^2. \tag{2.44}$$

If k is large enough - corresponding to a stiff spring - then we can approximate $F(\lambda) \approx \mathcal{W}(\xi)$.

The PMF is then

$$\mathcal{W}(\lambda_t) = \mathcal{W}(\lambda_0) - k_B T \ln\langle e^{-\frac{E_t}{k_B T}} \rangle \tag{2.45}$$

where E_t is the work done in moving from λ_0 to λ_t. The exponential average $\langle e^{-\frac{E_t}{k_B T}} \rangle$ is often approximated using a cumulant expansion.

Adaptive Biasing Force

The adaptive biasing force (ABF) method [175, 171, 176] for calculating the free energy profile along a reaction coordinate ξ applies a continuously changing biasing force to cancel the force acting along ξ. No average force then acts along ξ, so the evolution of the system is limited solely by its self-diffusion properties. In other words, the free energy landscape along ξ becomes smooth.

The average force acting along ξ, $\langle F_\xi \rangle_\xi$, is given by

$$\langle F_\xi \rangle_\xi = \frac{1}{P_\xi} \int \exp\left[-\mathcal{V}(\mathbf{x})/k_B T\right] \delta(\xi(\mathbf{x}) - \xi) \left(-\frac{\partial \mathcal{V}(\mathbf{x})}{\partial \xi} + k_B T \frac{\partial \ln |J|}{\partial \xi} \right) d\mathbf{x} \tag{2.46}$$

where $\mathcal{V}(\mathbf{r})$ is the potential energy function of the system, and $|J|$ is the determinant of the Jacobian for transforming from Cartesian to generalised coordinates. The probability density function P_ξ is given by

$$P_\xi = \frac{\int \delta(\xi(\mathbf{x}) - \xi) \exp(-\mathcal{H}(\mathbf{x}, \mathbf{p}_x)/k_B T)\, d\mathbf{x}\, d\mathbf{p}_x}{\int \exp(-\mathcal{H}(\mathbf{x}, \mathbf{p}_x)/k_B T)\, d\mathbf{x}\, d\mathbf{p}_x} \tag{2.47}$$

where $\mathcal{H}(\mathbf{x}, \mathbf{p}_x)$ is the Hamiltonian of the system. The PMF can be calculated from $\langle F_\xi \rangle_\xi$ by

$$\frac{d\mathcal{W}(\xi)}{d\xi} = -\langle F_\xi \rangle_\xi \tag{2.48}$$

For a derivation of equation 2.46 see references [171, 176].

2.5 Details of Simulations Presented in this Thesis

MD simulations in this thesis were performed using GROMACS 4.5.4 (www.gromacs.org) [177, 152] unless specified otherwise. Electrostatics interactions in all CG simulations were smoothly shifted to zero between 0 and 1.2 nm, and used a relative dielectric constant of 20. VDW interactions were smoothly shifted to zero between 0.9 and 1.2 nm. All simulations used constant NPT and periodic boundary conditions. Temperature and pressure coupling

used a Berendsen thermostat and barostat [153]. Helix, lipid and solvent (water plus ions) were separately temperature coupled with a coupling constant of 1 ps. Pressure coupling was semi-isotropic with a compressibility of 5×10^{-6} bar$^{-}1$. Na^{+} or Cl^{-} counterions were added to maintain electrical neutrality of the system. In PMF umbrella sampling simulations the distributions were unbiased using an implementation of WHAM by Alan Grossfield (membrane.urmc.rochester.edu) using a tolerance of 0.0001.

Chapter 3

Helix-Bilayer PMF Calculations

3.1 Introduction

It is often stated that hydrophobic effects drive the folding of most soluble proteins [178] and are a key component in explaining the folding and stability of membrane proteins [179, 180]. Many efforts have therefore been made to calculate the hydrophobicity of individual amino acid residues, mostly by attempting to calculate or measure their free energy of insertion into the centre of a bilayer [1, 181, 35, 182, 93], or a hydrophobic substitute [164, 37]. These calculations are relevant for protein function as well as folding. For example, the debate over the possibility of arginine residues crossing the centre of the bilayer as part of a voltage sensing function has been particularly vigorous, and centres around the energetic cost of transferring an arginine residue into the centre of a membrane [183, 184].

Hydrophobicity scales can be derived either from experiment or from theory. Examples of experimental free energy scales include Radzicka and Wolfenden's hydrophobicity scale [164], derived from the transfer of amino acid residues from water into cyclohexane; and the Wimley-White hydrophobicity scale [37], derived from the transfer of amino acid residues from water into octanol. The Wimley-White scale includes the contribution of the peptide bond as well as the side chain. The Kyte and Doolittle scale [185] combines data from the transfer of residues from water into vapour with the fraction of residues which are found inside the hydrophobic core of a representative set of proteins. A notable biological approach to finding a hydrophobicity scale was taken by Gunnar von Heijne and colleagues, using the translocon complex to insert model helices, flanked by glycosylation sites, into dog pancreatic ribosome membranes (as described in section 1.5.1). The distribution of singly *vs* doubly glycosylated sequences can be converted into an apparent free energy of insertion

[35, 36]. Theory-based hydrophobicity scales include continuum electrostatics calculations [181] and computational methods, typically taking the form of calculating PMF profiles for the transfer of individual residues or side chains across a lipid bilayer. This has been done at atomistic level [182] and using CG models, both MARTINI [93] and the Bond model [1].

Comparisons of all these scales suggest similar relative hydrophobicities; in other words, the scales broadly agree when identifying the most and least hydrophobic residues (see Table 3.1). However, the magnitudes of hydrophobicity measured by different methods are significantly different. The difference between the most and least hydrophobic residues is 83 kJ/mol for partitioning from water to cyclohexane; 24 kJ/mol for partitioning from water to octanol; 49 kJ/mol for continuum electrostatics calculations; and only 18 kJ/mol for translocon-mediated insertion.

It is therefore apparent that different methods provide different absolute estimates for free energies of insertion of residues into membranes. This is not unexpected, as each method calculates subtly different quantities. For instance, the computational methods explicitly measure the free energy profile for taking the residues across the membrane, whereas the translocon mediated insertion experiments measure the energy cost of inserting residues laterally into the membrane from the translocon pore. The latter quantity is unlikely to include the barrier involved in switching snorkelling (i.e. reaching up or down to interact with the bilayer headgroups) residues from one side of the bilayer to the other. Computational measures typically report a well depth ΔW whereas experimental methods report a free energy change ΔG, which are subtly different quantities (see section 3.2.2 for more detail). Other differences between scales may stem from the applicability of octanol and cyclohexane as mimics of the hydrophobicity of the membrane core.

Nevertheless, these hydrophobicity scales broadly agree on the order of hydrophobicities for each residue. But the differences may be instructive. A notable aspect of the hydrophobicity scales is that the insertion energies measured by translocon mediated insertion are significantly smaller in magnitude than those measured by most other methods, particularly computational approaches [184] (Table 3.1). One possible explanation for this is systematic errors in computational calculations (for example problems with force field parameters). However, this seems somewhat unlikely given the variety of different methods which have provided similar estimates [1, 182, 93] and the agreement with non-computational methods [181, 185]. A variety of other explanations have focused on lipid composition [186], protonation state [187], and density of protein in the membrane [188],

Table 3.1: Different hydrophobicity scales. Coarse-grained is abbreviated to CG. CG-B: Free energies of transfer of CG amino acid side-chain particles into DPPC bilayer from PMF calculations using the Bond coarse-grained model [1]. Glycine has no side-chain particle in the Bond model. CG-M: Free energies of transfer of CG amino acid side-chain particles from water into butane using the MARTINI coarse-grained model [93]. Neither glycine or alanine have side chain particles in the MARTINI model. R-W: Free energies of transfer of amino acid residues from water into cyclohexane [164]. Proline is not included because the proline sidechain does not capture its chemical structure when interacting with the protein backbone. AT: Free energies of transfer of small molecule analogues of amino acid sidechains from water into the centre of a DPPC bilayer from PMF calculations by atomistic simulations [182]. Glycine and proline have no suitable small molecule equivalents; histidine is not calculated owing to ambiguous protonation states. TR: apparent free energy of insertion of residue as obtained from translocon-mediated experiments [35]. EL: Free energies of transfer of amino acid side chains into the centre of a lipid bilayer as calculated using continuum electrostatic calculations [181]. Proline is not included owing to its helix-breaking properties. W-W: Free energies of transfer of whole amino acid residues from water into octanol [37]. K-D: A combination of free energies of transfer from water to vapour and the probability of finding the residue in the hydrophobic core of a test set of proteins [185]. Scales are ordered according to CG-B hydrophobicity.

	Hydrophobicity of residues (kJ/mol)							
Residue	CG-B	CG-M	R-W	AT	TR	EL	W-W	K-D
F	-34	-19	-12	-13	-1	-10	-7	-12
A	-20		-8	-9	0	-8	2	-8
I	-20	-22	-21	-22	-3	-14	-5	-19
L	-20	-22	-21	-15	-2	-15	-5	-16
V	-20	-20	-17	-14	-1	-11	-2	-18
W	-16	-12	-10	-5	1	1	-9	4
C	1	-9	-5	-3	-1	-9	0	-10
M	1	-9	-10	-4	0	-2	-3	-8
Y	7	1	1	7	3	4	-3	5
K	8	>25	23	50	11	23	11	16
S	16	11	14	16	4	4	2	3
T	16	11	11	14	2	-8	1	3
H	19	19	19		9	13	0	13
D	26	>25	36	80	15	30	15	15
E	26	>25	28	86	11	25	14	15
N	26	>25	28	24	9	17	4	15
Q	26	23	23	20	10	9	3	15
R	30	>25	62	60	11	34	8	19
G			-4		3	-7	5	2
P		-20			9		1	7

with limited success.

An alternative explanation for the disparity between *in vivo* and theoretical estimates of hydrophobicity lies in the high level of complexity in the cellular environment in which the experiment occurs. It is possible that the relationship between insertion efficiencies from translocon-mediated experiments and actual free energies of insertion into membranes may be more complicated than previously realised. The precise meaning of the ΔG_{app} values measured in the experiments depends on the mechanism of the translocon and on which steps in the insertion process occur at equilibrium. If the translocon simply acts as a passive transporter, reducing the barrier preventing the helix entering the membrane, the ΔG_{app} would indeed reflect the free energy difference between the helices in some 'non-inserted' state and those in the 'inserted' state. Which steps occur at equilibrium, if any, would determine the nature of the 'non-inserted' state. However, if the translocon plays a more active role, then the ΔG_{app} would instead reflect some more complex pathway, perhaps including some conformational change in the translocon. Comparison of free energies of transfer from water to the membrane with those obtained from translocon-based assays may in principle therefore provide information about the function of the translocon. However, identifying the method which provides the best estimate of the absolute free energy of insertion is not straightforward.

Attempts have been made to calculate PMF profiles for the insertion of helices into a bilayer [1, 189]. Most - if not all - examples to date have involved simple model helices. These have been successful in elucidating some of the physicochemical principles of insertion and helix stability in a bilayer but helices in membrane proteins typically show a much higher degree of sequence variation than these simple model helices. I therefore calculated the PMF of 'biological' helices (those with sequences taken from membrane proteins) crossing a lipid bilayer. This introduces another layer of complexity compared with model helices, so both tests the applicability of a widely used technique to more complex systems and brings us closer to understanding the energetics of *in vivo* helix insertion.

3.1.1 CFTR

The helices chosen to investigate came from the cystic fibrosis transport regulator (CFTR). This is a chloride channel which has a key role in regulating salt and water homeostasis [190]. The insertion properties of these helices have been very well characterised by translocon-mediated insertion experiments [191]; in particular, Enquist and colleagues mea-

sured the insertion efficiencies of 19-residue scans along the 39 residue sequence surrounding the second transmembrane helix (M2) from CFTR [191]. This 39 residue sequence provided the basis for the helices chosen to investigate insertion into membranes. The full 39-residue sequence (termed M2-39) was used, together with the highest-inserting 19 residue section according to the translocon-mediated experiments (M2-19) and a medium-length segment (M2-29). These helices are shown in Figure 3.1 in blue, red and green respectively. This system provides a quantitative understanding of how *in vivo* insertion behaviour varies as a function of sequence, and so is a suitable system to investigate discrepancies between computational and *in vivo* measures of insertion. These differences may be able to shed some light on the mechanism of the translocon.

3.2 Methods

3.2.1 Theory

The potential of mean force (PMF) $\mathcal{W}(\xi)$ along some chosen coordinate ξ can be written as

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right] \tag{3.1}$$

with ξ^* an arbitrary value on the chosen coordinate. $\langle \rho(\xi) \rangle$ is the average distribution function, defined as

$$\langle \rho(\xi) \rangle = \frac{\int \delta(\xi'(\mathbf{R}) - \xi) e^{-U(\mathbf{R})/k_B T} d\mathbf{R}}{\int e^{-U(\mathbf{R})/k_B T} d\mathbf{R}} \tag{3.2}$$

For the helix-bilayer PMF calculations in this chapter, $U(\mathbf{R})$ is the total energy of the simulated system. The reaction coordinate ξ is the separation between the helix and bilayer centre of masses, projected on to the axis lying perpendicular to the bilayer (z). $\xi'(\mathbf{R})$ represents the position of the centre of mass of the helix along the reaction coordinate in a given simulation. The integral $\int d\mathbf{R}$ is over all non-sampled degrees of freedom in the system.

3.2.2 Relationship between PMF and experimental results

The quantity measured in translocon mediated experiments is ΔG_{app}, defined as

$$\Delta G_{app} = -RT \ln \left(\frac{f_{1g}}{f_{2g}} \right) \tag{3.3}$$

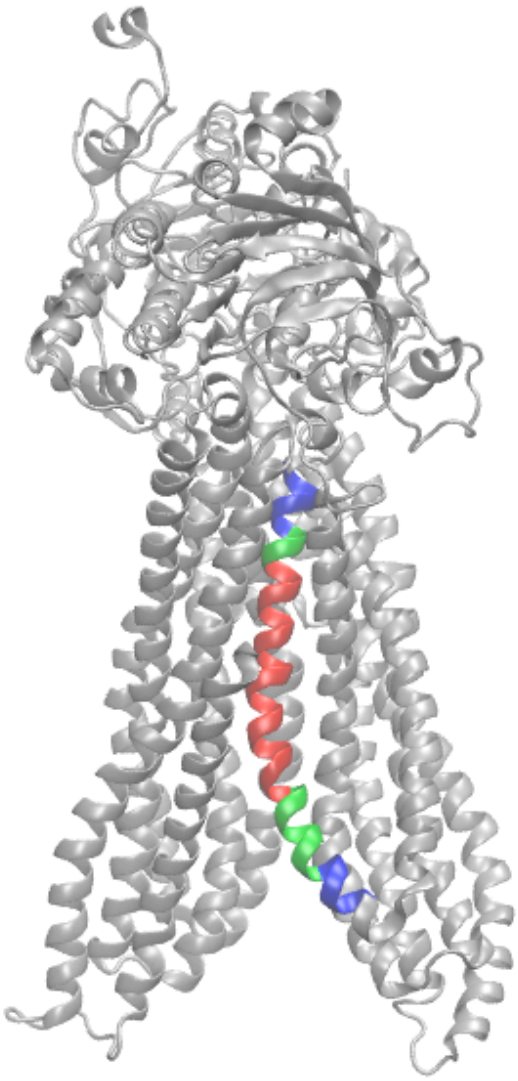


Figure 3.1: CFTR including M2-19 (red), M2-29 (edges in green) and M2-39 (edges in blue). The structure comes from a homology model of Sav1866 refined by molecular dynamics [192].

where f_{1g} and f_{2g} are the fraction of singly and doubly glycosylated helices respectively [35]. The equivalent quantity from the PMF $\mathcal{W}(\xi)$ is

$$\Delta G_{pmf} = -RT \ln \left(\frac{Q_{ins}}{Q_{nins}} \right) \tag{3.4}$$

with Q_{ins} and Q_{nins} the partition functions for the inserted (transmembrane) and not-inserted (in water and interfacial) states respectively [193], defined as

$$Q_{ins} = \int_{inserted} \exp \left(\frac{-\mathcal{W}(\xi)}{k_B T} \right) d\xi \tag{3.5}$$

$$Q_{nins} = \int_{not\ inserted} \exp \left(\frac{-\mathcal{W}(\xi)}{k_B T} \right) d\xi \tag{3.6}$$

Defining the value of ξ between the 'inserted' and 'not inserted' states (ξ_{trans}) is not always straightforward. For PMF profiles with a significant interfacial (IF) energy well, the transition point is chosen to be where at the highest energy point between the IF and TM wells. For profiles without a significant IF well, the value of the reaction coordinate where the helix tilt switches from $\sim$90° to $\sim$0° is used. ΔG_{pmf} is sensitive to the choice of ξ_{trans} for many profiles.

For the problem of a ligand binding to a protein, the free energy of binding ΔG_{bind} as measured by simulations includes an implicit volume-dependent contribution ΔG_V [193]. This ΔG_V represents the free energy cost of taking the ligand from a standard state volume V° to the volume of water sampled by the ligand in the unbound state $V_{unbound}$, and is given by

$$\Delta G_V = -RT \ln \left(\frac{V_{unbound}}{V^\circ} \right) \tag{3.7}$$

By convention V° is chosen to be 1661 Å^3, corresponding to a ligand concentration of 1 M. It is unclear that a similar volume correction factor needs to be used for the helix-bilayer PMF calculations in this chapter. Provided the helices sample a similar volume of water in each PMF calculation - which they do, owing to similar box sizes - and the calculation is performed over the same range of z values, then the values of ΔG_{pmf} are comparable between different PMFs. As the helices have a 'binding site' which is a lipid bilayer extending across the box in x and y dimensions, the comparison with a limited binding site in one protein is not valid. The standard state in the translocon-mediated experiments is also unclear, so no attempt was made to find a value of ΔG_V for these PMF calculations.

It may also be useful to use the depth of the PMF to compare with particular quantities.

Table 3.2: PMF calculations performed. Orientation refers to the angle the helix makes to the axis perpendicular to the bilayer. Length is the length of the simulation windows, and number refers to the number of windows in the PMF calculation. 'Helical' refers to the type of secondary structure restraints applied.

Helix	Lipid	Orientation (°)	Helical	Length (ns)	Number
M2-19	DPPC	0	All	40	120
M2-29	DPPC	45	All	60	140
M2-39	DPPC	0	All	80	150
M2-39	DLPC	0	All	100	150
M2-39	DHPC	0	All	60	150
M2-39	DPPC	90	All	60	150
M2-39	DPPC	0	Partial	60	150
M8-39	DPPC	0	All	80	150
M8-39	DPPC	90	All	60	150

The various useful depths are

$$\Delta\mathcal{W}_{wc} = \mathcal{W}(\xi_{core}) - \mathcal{W}(\xi_{water}) \tag{3.8}$$

$$\Delta\mathcal{W}_{wi} = \mathcal{W}(\xi_{interface}) - \mathcal{W}(\xi_{water}) \tag{3.9}$$

$$\Delta\mathcal{W}_{ic} = \mathcal{W}(\xi_{core}) - \mathcal{W}(\xi_{interface}) \tag{3.10}$$

where $\mathcal{W}(\xi_{area})$ is the minimum value of the PMF evaluated when the helix is in $area$ - either water, the interface or the membrane centre.

3.2.3 PMF Calculations

The PMF calculations performed are summarised in Table 3.2. These were designed to investigate the effect of a number of factors on the stability of helices in a bilayer: helix length, bilayer width, the flexibility of the edges of a helix, and the hydrophobicity of the helix. Simulations were also performed with the helices starting parallel to the bilayer, rather than perpendicular, to investigate whether the initial orientation of the helix could bias the conformation it adopted in the simulations. Each PMF calculation took approximately 40,000 CPU-hours, depending on the size of the system.

3.2.4 Simulation Details

Atomistic helices were generated from the required sequence using the build function of the Pymol program [194]. This produces helices with ϕ and ψ angles of $\sim$-56° and $\sim$-47°

respectively. Local code then generated a coarse-grained representation. The helix was aligned as required, then added to a simulation box containing a pre-equilibrated CG bilayer, with the distance between the centre of mass of the helix and the centre of mass of the bilayer being the value of the reaction coordinate sampled by the particular window. Lipid molecules which overlapped with the helix were removed. The system was then subjected to 1000 iterations of steepest descent energy minimisation, with the helix restrained. Water molecules were then added, and the system similarly energy minimised, with both the helix and lipids restrained. Sufficient counter-ions were then added to ensure that the system was electrically neutral. A 3 ns equilibration simulation was then performed with helix and lipids restrained; this ensured that all water particles were expelled from the bilayer. The production simulation was then performed, restraining the centre of mass of the helix relative to the centre of mass of the bilayer with a harmonic restraint, with force constant of 1000 kJ/mol/nm^2. The displacements of the centre of mass of the helix from the restraint point were recorded. These were then unbiased using Alan Grossfield's implementation of WHAM [195]. The number of bins used was 200 with a tolerance of 0.0001. A summary of all PMF calculations performed is in Table 3.2. Helix tilt angles were defined as the angle between the vector joining the centre of mass of the three CA particles at the ends of the helices, and the vector lying perpendicular to the plane of the bilayer.

Some of the PMF profiles showed a different energy level for the helix in bulk water on either side of the bilayer. This is not physically reasonable, and is probably a result of kinetic trapping in some windows combined with the restriction on WHAM which ensures that the profile remains continuous. I therefore applied a 'clamping' restraint to each profile, requiring the ends of the profile - where the helix is in bulk water - to be 0 kJ/mol.

The first 10 ns of each simulation were not used to calculate PMF profiles, as they were treated as equilibration time. The profiles shown are the average of the profiles calculated over the first and second halves of the remaining simulation time, except where otherwise indicated. The error bars shown are the difference between the two profiles. These should be interpreted as a measure of sampling effectiveness given possible kinetic trapping biased by starting position occurring in some windows. They are not a measure of the statistical error encompassing all possible states of the system.

All simulations in this chapter were performed using Gromacs version 3.3.3 [177]. The temperatures of the helix, lipids and solvent were each coupled separately using the Berendsen algorithm [153] at 323 K. The time step was 40 fs. The pressure was semi-isotropically

Table 3.3: Sequences of helices

Helix	Sequence
M2-19	IAIYLGIGLCLLFIVRTLL
M2-29	ERSIAIYLGIGLCLLFIVRTLLLHPAIFG
M2-39	PDNKEERSIAIYLGIGLCLLFIVRTLLLHPAIFGLHHIG

coupled using the Berendsen algorithm at 1 bar with τ_P = 1 ps and a compressibility of 5 $\times$ 10^{-6} bar^{-1}.

3.3 Results

3.3.1 Helix Length

The sequence chosen for study was taken from the second transmembrane helix from CFTR. The insertion efficiencies of 19-residue windows along the 39 residue sequence surrounding this helix have been measured by translocon-mediated insertion experiments [191]. These results clearly showed that the 19-residue sequences inserted into the membrane more efficiently the closer they were to the native sequence (with one exception, M2-19(-1), which inserted more efficiently than the native helix, M2[1]). This was a suitable system to assess the similarity between insertion free energies as measured by *in vivo* experiments and those measured computationally, as it is possible to compare both quantitative ΔG values and qualitative behaviour (insertion dependence on sequence). I therefore performed PMF calculations on M2-39 to compare qualitative behaviour with experiment; M2-19 to quantitatively compare the insertion free energies obtained with the two methods applied to identical sequences; and M2-29 to check the dependence of the PMFs on helix length. The respective sequences are shown in Table 3.3. Pictures of the helices, with residues coloured according to chemical properties, are shown in Figure 3.2.

The PMF profiles are shown in Figure 3.3. Table 3.4 shows the dependence of ΔG_{pmf} on helix length. The insertion of helices M2-19, M2-29 and M2-39 becomes much less favourable as the length of the helix shortens, with values of ΔG_{pmf} of 24, -49 and -89 kJ/mol for the three helices respectively. M2-19 is too short to fully span the bilayer (see Figure 3.4), so

[1]M2 refers to the 19 residue physiological sequence; M2-39 refers to the 39 residue sequence surrounding M2, for which insertion efficiencies have been obtained from translocon-mediated insertion experiments; M2-29 refers to the 29 residue sequence surrounding M2; and M2-19 refers to the 19 residue sequence which translocon mediated experiment suggested had the highest insertion efficiency. M2-19 corresponds to the M2-19(-1) sequence in the notation of Enquist et al., and corresponds to the 19 residue sequence which is shifted by 1 residue to the left from M2.

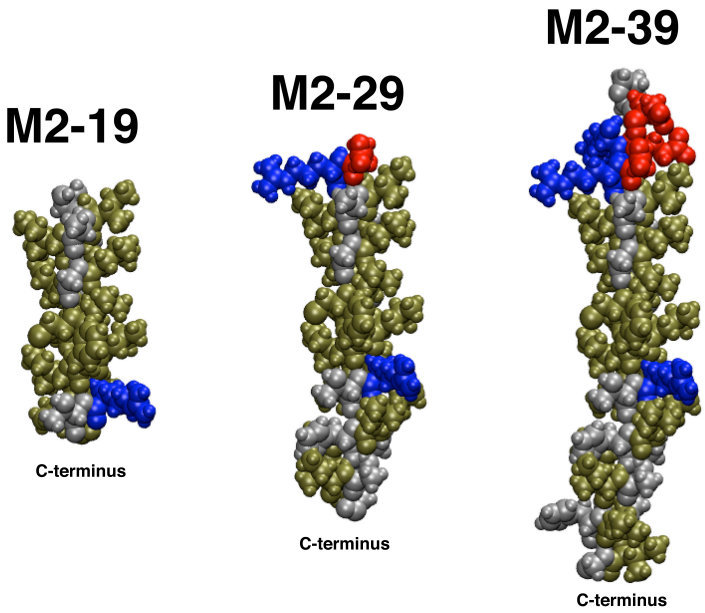


Figure 3.2: Helices of different lengths taken from the sequence of M2 from CFTR. Positively charged residues (Arg and Lys) are blue; negatively charged residues (Asp and Glu) are red; hydrophobic residues (Val, Ile, Leu, Met, Phe, Cys) are tan; all others are grey.

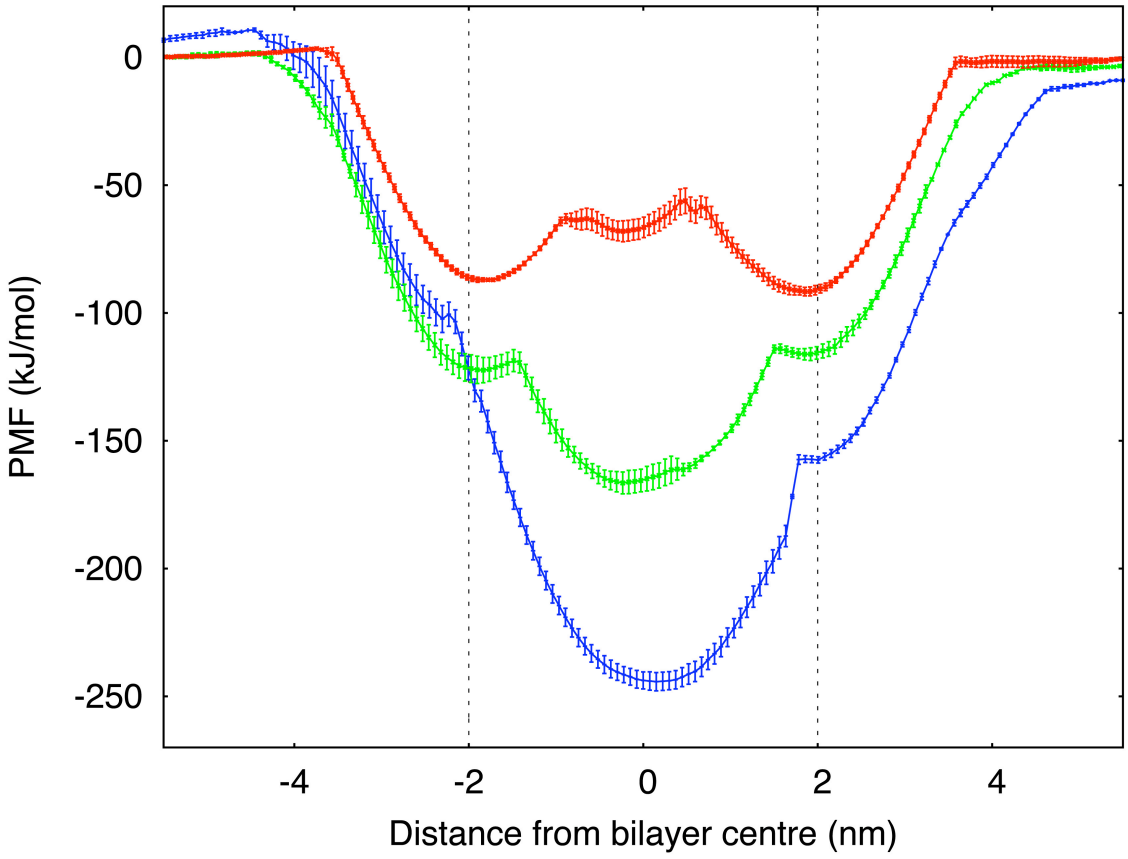


Figure 3.3: PMF profiles for the transfer of M2-19 (red), M2-29 (green) and M2-39 (blue) across a DPPC bilayer. Approximate edges of the bilayer are shown in vertical black dashed lines.

Table 3.4: Free energies obtained from PMF profiles. M2-39$_s$ indicates a simulation where the initial position of the helix was parallel with the bilayer. M2-39$_p$ indicates a sequence where only the central 19 residues were restrained as an ideal helix. ξ_{trans} indicates the point along the reaction coordinate chosen as the boundary between the 'inserted' and 'not-inserted' states.

Helix	Lipid	ΔG_{pmf} (kJ/mol)	$\pm\xi_{trans}$ (nm)
M2-19	DPPC	+24	1
M2-29	DPPC	-49	1.5
M2-39	DPPC	-89	1.9
M2-39	DLPC	-45	1.5
M2-39	DHPC	-34	1.1
M2-39$_s$	DPPC	-10	1.9
M2-39$_p$	DPPC	-82	1.9
M8-39	DPPC	+45	1.1
M8-39$_s$	DPPC	+41	1.2

for the ends to stay hydrated the bilayer has to deform, incurring a significant energetic penalty. The interfacial (IF) state, where hydrophobic residues can mostly interact with hydrophobic lipid tails and the hydrophilic end residues can interact with polar headgroups and water molecules (see Figure 3.5), is therefore stable for M2-19; the distortion in the lipid bilayer means that the transmembrane (TM) state is less stable than the IF state. The average phosphate-phosphate distance as a function of distance from the helix is also shown for each of the helices in Figure 3.6. Some distortion is seen in all simulations, but both M2-29 and M2-39 cause less distortion than M2-19 when spanning the bilayer. For these helices the TM state is more favourable than the IF state as all hydrophobic residues are shielded from polar headgroups and water particles. In particular, M2-39 can adopt tilt angles and kink (about Pro^{30}, near the C-terminus) to allow hydrophobic residues to stay in the membrane core and polar and charged residues to interact with polar headgroups (see Figure 3.7). This ability for the helix to adapt to the bilayer allows it to maximise favourable interactions between the helix residues and the bilayer, while minimising energetically costly distortions to the bilayer, and explains why the TM position is much more favourable than the IF position. The length-dependence of helix insertion efficiencies has also been observed in translocon-mediated insertion experiments, supported by atomistic molecular dynamics simulations [196].

In order for the PMF calculations to be accurate, all degrees of freedom perpendicular to the reaction coordinate should be equilibrated. For these helix-bilayer PMFs, an important degree of freedom is the rotation of the helix with respect to the plane of the bilayer. The

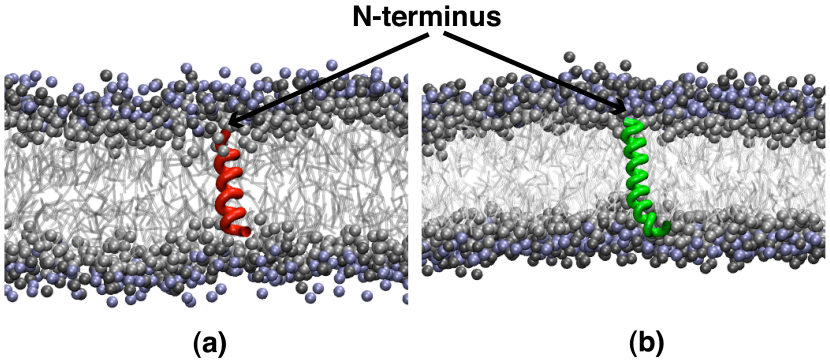


Figure 3.4: Bilayer distortion caused by (a) M2-19, red, and (b) M2-29, green, in a transmembrane (TM) state (0 nm between helix centres of mass and bilayer centre of mass). Lipid headgroup particles are coloured grey and iceblue, while water particles have been omitted for clarity.

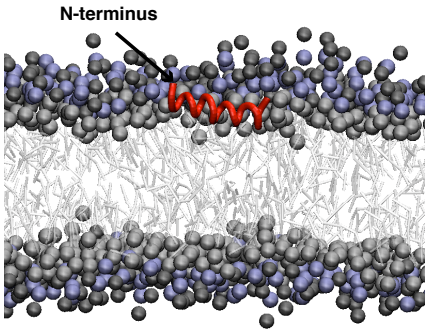


Figure 3.5: M2-19 in an interfacial (IF) state (2 nm separation between helix and bilayer centres of mass). Lipid headgroup particles are coloured grey and iceblue, while water particles have been omitted for clarity.

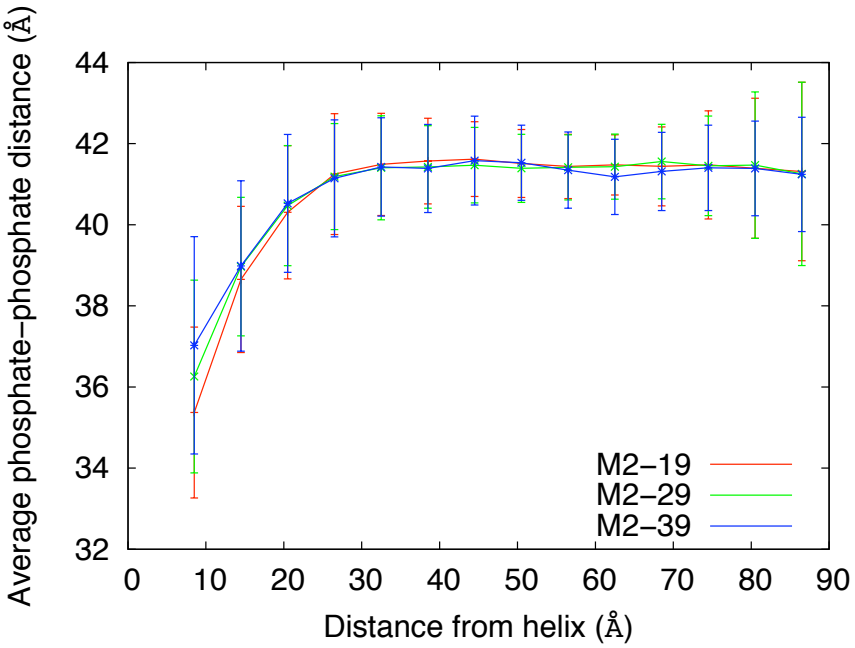


Figure 3.6: Average phosphate-phosphate distances for DPPC taken from simulations with M2-19 (red), M2-29 (green) and M2-39 (blue) in a TM position. Error bars represent the standard deviation of the phosphate-phosphate distances across the simulations.

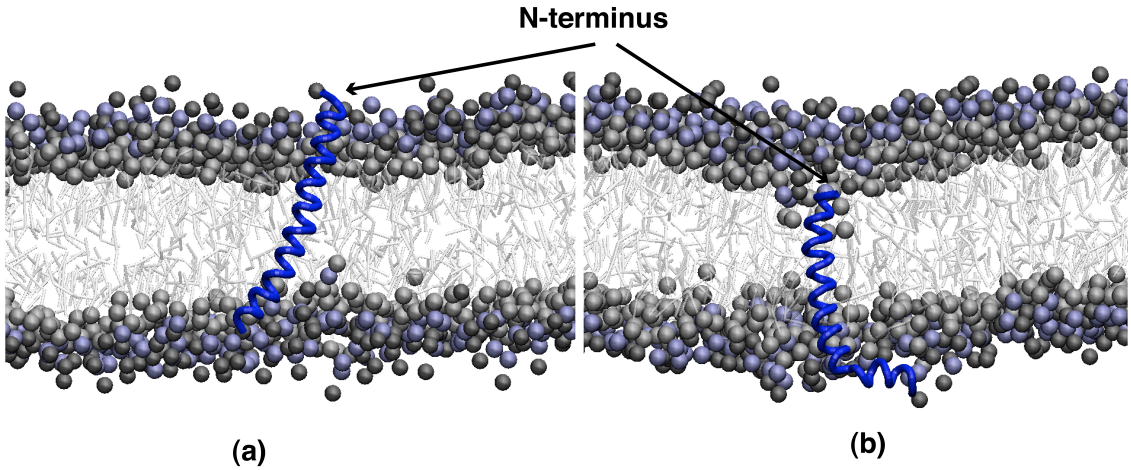


Figure 3.7: M2-39 in TM orientations, showing the ability to tilt and kink in order to match hydrophobic widths to the bilayer. (a) 0.8 nm separation between helix and bilayer centres of mass. (b) -1.8 nm separation between helix and bilayer centres of mass.

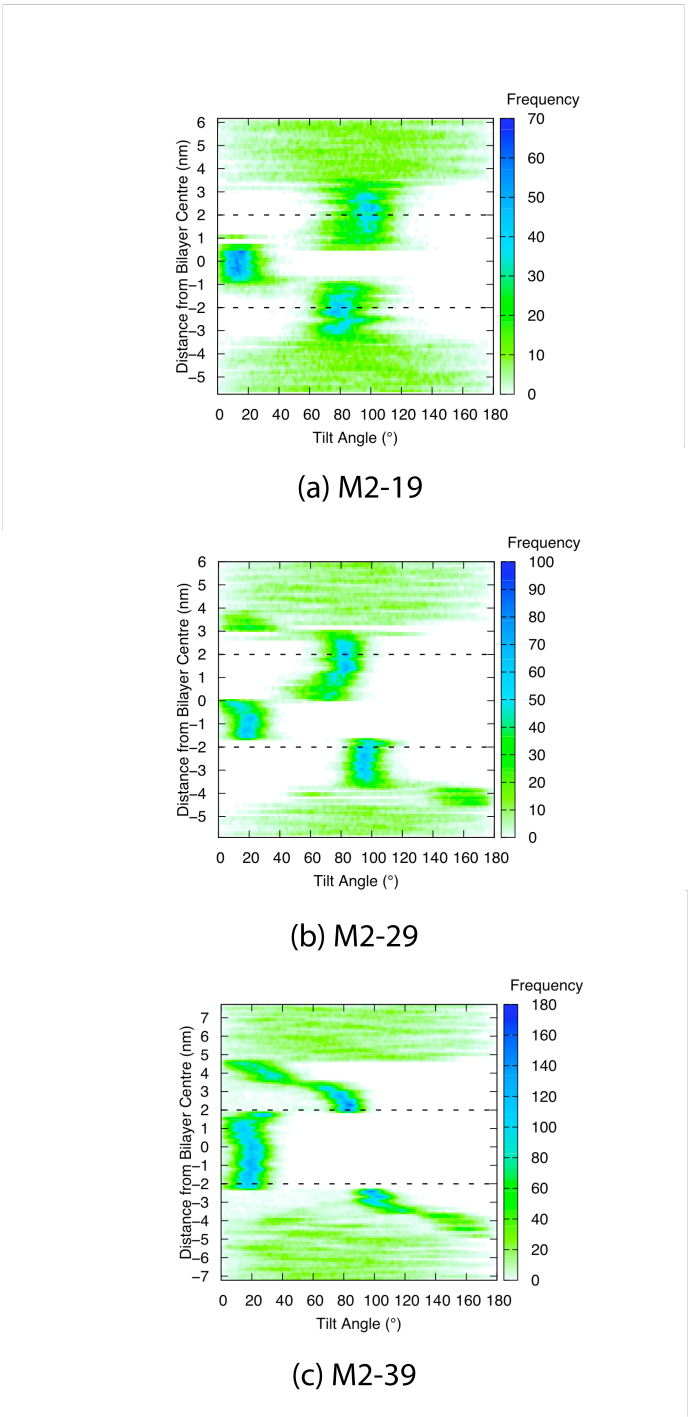


Figure 3.8: The distribution of helix tilt angles from the M2-19, M2-29 and M2-39 simulations. Zero degrees is perpendicular to the plane of the membrane. The dashed black lines show the approximate edges of the DPPC bilayer.

frequency of the helix tilt angle is shown as a function of reaction coordinate for each helix in Figure 3.8. These provide an informative view of the behaviour of the helices as their position relative to the bilayer changes. For M2-19, all tilt angles are equally sampled in water (when more than $\sim$3 nm from the centre of the bilayer), showing that the helix is tumbling freely. For the region where the centre of mass of M2-19 is within $\sim$1 nm of the edges of the bilayer the helix clearly adopts an IF position, corresponding to a tilt angle of $\sim$90°. When the centre of mass of the helix is within $\sim$1 nm of the centre of the bilayer it adopts a TM position, with a tilt angle of $\sim$0°. The behaviour of the helix is symmetric with respect to the bilayer, indicating that it does not depend on the starting position as strongly as longer helices.

In contrast, the behaviour of M2-29 is noticeably non-symmetric with respect to the bilayer. When the starting position of the simulation is with the N-terminus pointing towards the bilayer, the helix adopts an IF position when $\sim$3.5-1.8 nm from the centre of the bilayer, and adopts a TM position when $\sim$1.8-0 nm from the bilayer centre. Curiously, when the starting position is with the N-terminal end of the helix pointing away from the bilayer, the IF state seems much more favoured. The helix adopts this state when as little as $\sim$0.5 nm away from the centre of the membrane. It seems that the charged Arg^{7} and Glu^{6} at the N-terminal end anchor this end at the interface; meanwhile Arg^{24} (closer to the C-terminal end) snorkels up towards the interface, drawing the C-terminal end of the helix up into an IF position (see Figure 3.11). It seems likely that if the helix were to start in a vertical position with respect to the bilayer, as with the other calculations, then it would have adopted a TM position. The helix adopting an IF position while its centre of mass is restrained around the centre of the bilayer is an example of kinetic trapping caused by the choice of starting structure.

M2-39 exhibits largely symmetric behaviour, characterised by adopting an IF position when the centre of the helix is within $\sim$2 nm of the centre of the bilayer. However, when the N-terminal end is facing the bilayer centre and the helix is $\sim$3-5 nm away from the bilayer centre, the hydrophobic Ile^{32}, Phe^{33} and Leu^{35} residues maintain an interaction with the headgroups of the bilayer, resulting in restricted tilt angles of $\sim$10-80° depending on the distance away from the bilayer (see Figure 3.9). The helix adopts a fully IF position in only a few simulations.

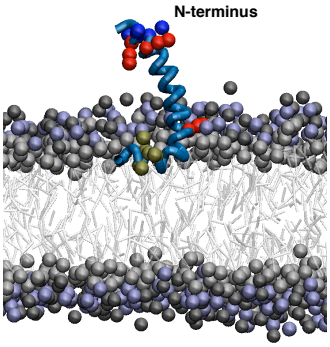


Figure 3.9: M2-39 with a hydrophobic end of the helix staying in the bilayer while the rest of the helix is in water. This snapshot comes from a simulation where a restraint is applied at 3.7 nm separation between the helix and bilayer centres of mass.

Asymmetries in the M2-29 Profile

The PMF profile for M2-29 calculated directly from simulation data exhibits some unphysical attributes - in particular there appears to be a difference of $\sim$60 kJ/mol between IF states above and below the bilayer (see Figure 3.10). In an IF state hydrophobic residues would try to interact with the lipid tails and polar residues would try to interact with lipid headgroups or water. It is reasonable to expect that IF states on different sides of the bilayer would include similar interactions, so should be similar in energy.

This difference between the energies of the IF states is a result of M2-29 adopting an IF state when its centre of mass is 0-2.5 nm away from the bilayer centre. On the other side of the bilayer, in contrast, it adopts a TM position when -1.5-0 nm away from the bilayer centre. This is caused by the starting structure of the simulations; the helix starts at an angle of 45° to z, and contains a charged arginine at position 24. This arginine attempts to snorkel towards the bilayer headgroups. When the helix starts with the N-terminus pointing towards the bilayer centre (Figure 3.11 (a)), Arg^{24} can snorkel down towards the C-terminus end of the helix, thus drawing that end down so the helix adopts a TM position. When the helix starts with the N-terminus pointing away from the bilayer centre, the Arg^{24} snorkels instead up towards the N-terminus of the helix (Figure 3.11 (b)), thus drawing it up into an IF position. This has the effect of biasing the calculated probability distribution function and so produces an unphysical PMF profile. This is similar to behaviour that has been observed in CGMD simulations of the GWALP peptide [197], where positioning an arginine residue in position 12 produces a different peptide orientation from positioning the arginine in position 14, due to the arginine snorkelling either in the N-terminal or C-terminal direction [138]. This behaviour is consistent with solid-state NMR experiments.

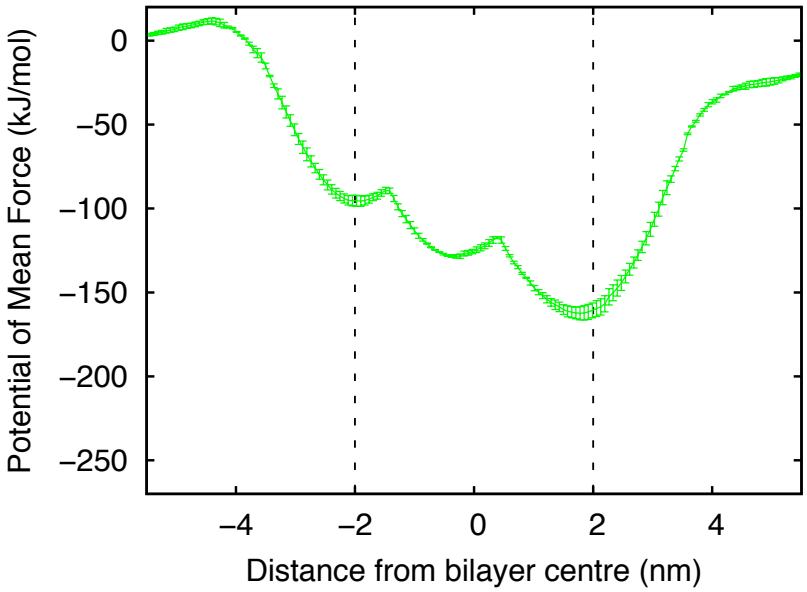


Figure 3.10: PMF profile for translating M2-29 across a DPPC bilayer, as obtained directly from the simulations.

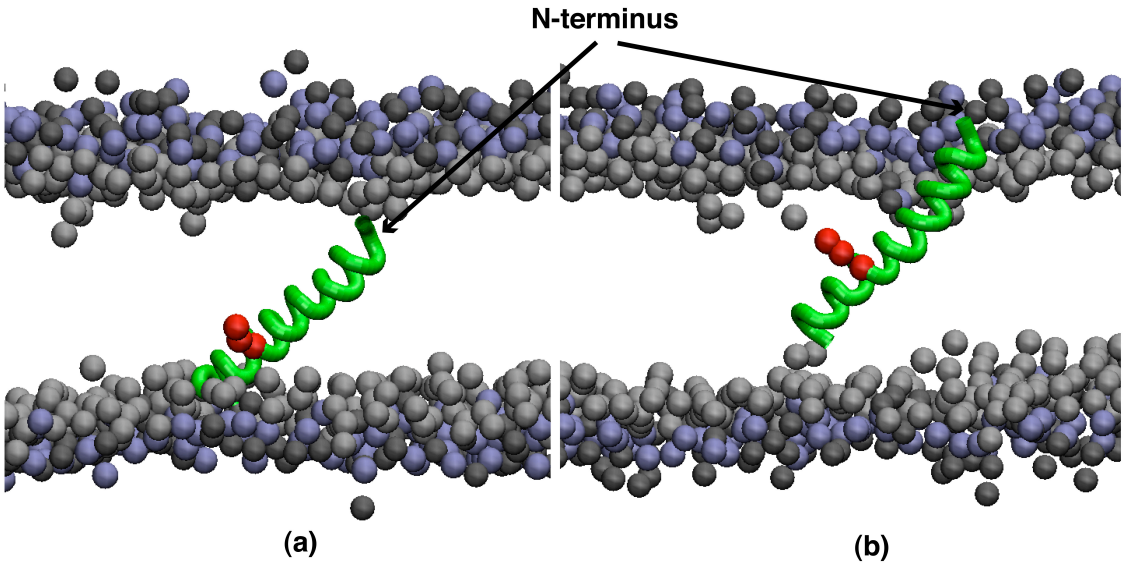


Figure 3.11: M2-29 in its starting orientations with the centre of mass of the helix equivalent distances above and below the bilayer. On the left, with the N-terminus closest to the centre of the bilayer, Arg^{24} (shown in red) will snorkel down to the bilayer headgroup region, dragging the helix into a TM position. On the right, with the N terminus further from the centre of the bilayer, Arg^{24} will snorkel up to the bilayer headgroups, drawing the helix up into an IF position.

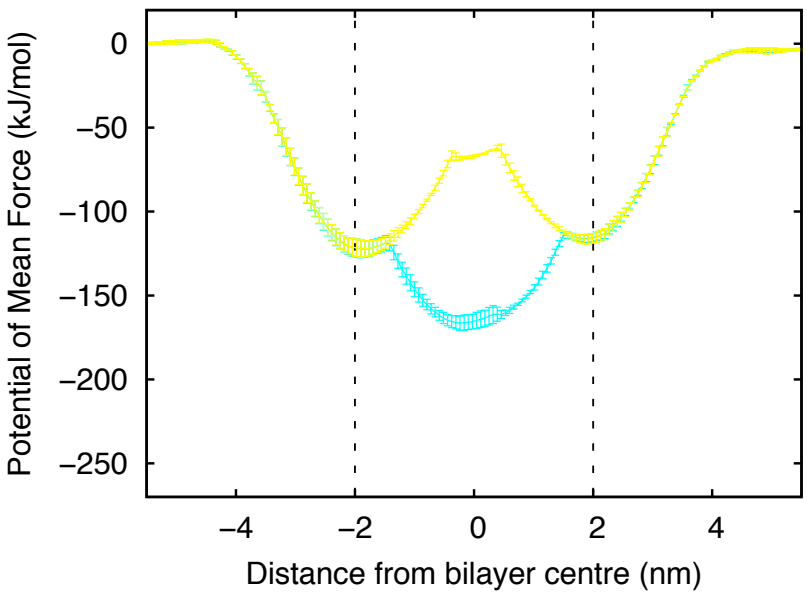


Figure 3.12: PMF profiles for translating M2-29 across a DPPC bilayer. In these profiles the displacements of the helix away from the centre of the restraining potential were estimated by inverting the distribution from the 'opposite window' for 13 windows. The 'opposite window' was the equivalent window on other side of the bilayer. The yellow profile shows where distributions from the negative reaction coordinate were replaced; the cyan profile shows where distributions from the positive reaction coordinate were replaced. They are identical apart from in the centre of the membrane. See section 3.3.1 for further details.

To address the problem of IF states with different energies I discarded data from the simulations where a kinetic trapping was evident. Instead I used inverted data from equivalent simulations on the other side of the bilayer. I identified 17 windows where the behaviour of the helix was qualitatively different depending on which side of the bilayer the restraint was applied, and used the inverted displacements from the equivalent window on the other side of the bilayer to calculate the PMF. It was not immediately obvious which windows should be discarded, so I calculated two profiles: replacing first one set of data from one side of bilayer, then the other. The two profiles can be thought of as showing two different pathways through the free energy landscape. The resulting profiles in Figure 3.12 clearly show that the choice of data used has a profound effect on the PMF.

The suggested free energy landscape for M2-29 is sketched in Figure 3.13. This shows a deep TM well, and two IF wells near the edges of the bilayer similar in energy. Suggested energy barriers between the IF and TM states (shown in red) are not symmetric with respect to the bilayer as they are dependent on the snorkelling direction of Arg[24]. Figure 3.13

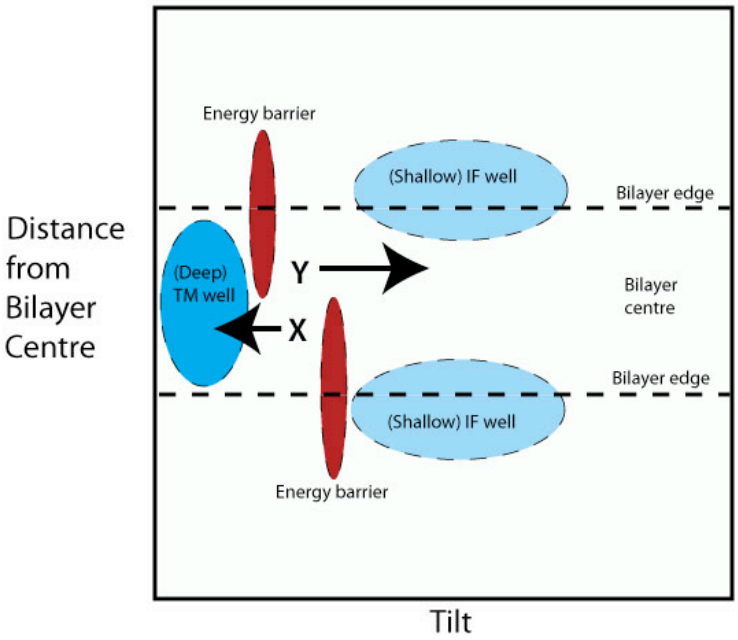


Figure 3.13: Schematic of suggested energy landscape for M2-29 as a function of helix distance from bilayer and helix tilt with respect to bilayer normal. A helix starting at an angle of 45° with its N-terminal end pointing towards to the bilayer centre (position X) will move towards the TM energy well, while a helix starting with its N-terminus pointing in the other direction (position Y) will move towards an IF position.

shows how helices starting at 45° to the bilayer normal will be in either side of these barriers depending on direction of the N-terminus of the helix, and so the helices will end up in different states depending on their starting position.

I considered that the distribution resulting in a well in the TM region (turquoise profile in Figure 3.12), rather than a barrier (yellow profile in Figure 3.12), reflected the reaction coordinate of most physical interest for this system. This is for the following reasons. 1) It seems more likely that the helix is kinetically trapped in the interfacial rather than transmembrane states, considering both the hydrophobic nature of the residues at the centre of the helix and the behaviour of the same residues in M2-19 and M2-39 simulations. Hence using data from simulations where the helix adopted a TM rather than IF state is more appropriate. 2) This results in a profile with a much lower energy state in the centre of the membrane, as shown in Figure 3.12. This suggests that this choice reflects the helix sampling a lower energy state when in the transmembrane state than in the interfacial state. 3) It results in a profile which fits the pattern of length dependence compared with M2-19 and M2-39 profiles.

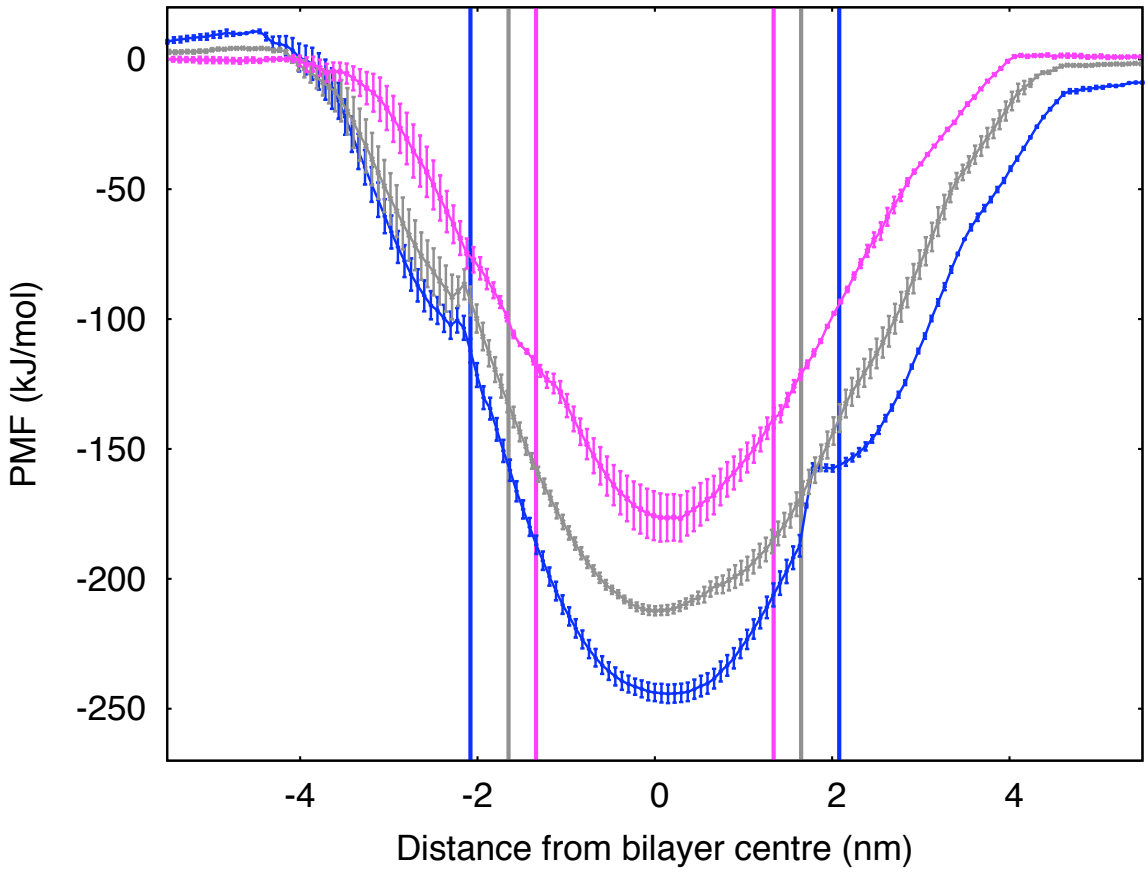


Figure 3.14: PMF profiles for the transfer of M2-39 across a DPPC bilayer (blue), a DLPC bilayer (grey) and a DHPC bilayer (purple). Approximate edges of the bilayers are shown in vertical lines in the corresponding colour.

3.3.2 Bilayer Width

The constriction point of the channel at the centre of the translocon is formed by the pore ring of six hydrophobic residues (see Figure 1.4) [22]. *In vivo* the helices therefore pass through a thin hydrophobic region before being transferred either into or across the membrane. Mutations of these residues have been shown to affect insertion efficiencies of helices [198]. This functional significance of the pore ring provided the motivation to investigate the effect of hydrophobic width on ΔG_{pmf}. I performed a PMF calculation of M2-39 crossing a lipid bilayer with three different hydrophobic widths. These bilayers consisted of lipid types DHPC, DLPC and DPPC, which have identical headgroups but different tail lengths (see Figure 3.15). The bilayers they form have phosphate-phosphate widths of ∼2.7 nm, ∼ 3.3 and ∼4.2 nm respectively. The comparable widths of the hydrophobic core are ∼1.4 nm, ∼2.0 nm and ∼2.9 nm respectively. These compare with a phosphate-phosphate distance for the protein-depleted rat hepatocyte endoplasmic reticulum membrane of 3.9 nm and the apical plasma membrane of 4.2 nm [199].

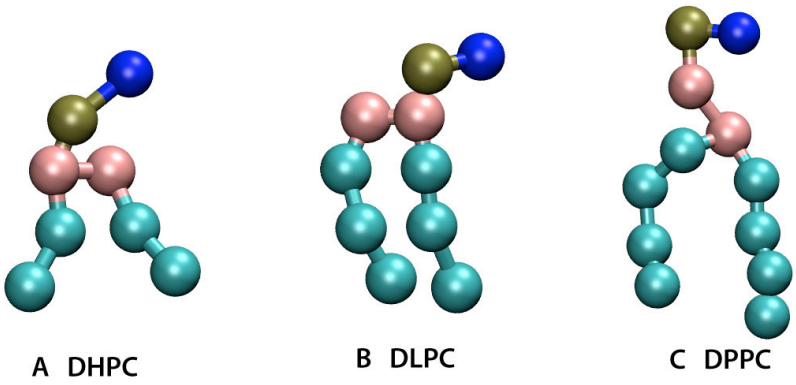


Figure 3.15: CG representation of the different lipids used. The phosphate particle is blue, choline brown, glycerols pink and carbon tails cyan. A) DHPC. B) DLPC. C) DPPC.

Comparisons of the DPPC, DLPC and DHPC M2-39 PMF profiles (Figure 3.14) show that values of ΔG_{pmf} are significantly reduced when the hydrophobic core of the bilayer is reduced. This is reasonable, as a thinner bilayer core means fewer favourable interactions between lipid tails and hydrophobic helix residues. The helix also has to tilt or kink more in order to maximise the possible favourable interactions.

The values of ΔG_{pmf} are -34, -45 and -89 kJ/mol for the DHPC, DLPC and DPPC simulations respectively. The correlation with hydrophobic width is shown in Figure 3.16. A straight line of best fit is shown for guidance, although it is impossible to be certain of a linear relationship from only three data points. The line of best fit has a gradient of -37, suggesting that reducing the hydrophobic width of the bilayer by 1 nm would increase the ΔG_{pmf} by 37 kJ/mol.

It is apparent that the thinner hydrophobic region results in a significantly smaller ΔG_{pmf} for M2-39. This suggests that the hydrophobic pore ring in the centre of the translocon channel will have some effect on the insertion efficiencies of helices *in vivo*, and therefore will affect the measured ΔG_{app} values obtained from translocon-mediated insertion experiments.

There are no metastable energy wells apparent in the PMF profile for the DLPC or DHPC calculations, as the helix adopted an IF position only in a few windows. This is probably because for the thinner bilayers the starting point of most windows includes the helix spanning both headgroup regions. The charged and aromatic residues at the ends of the helices therefore help to anchor the helix in the TM state. In contrast, in the DPPC simulations, there are more windows where only one end of the helix interacts with the headgroup region - in

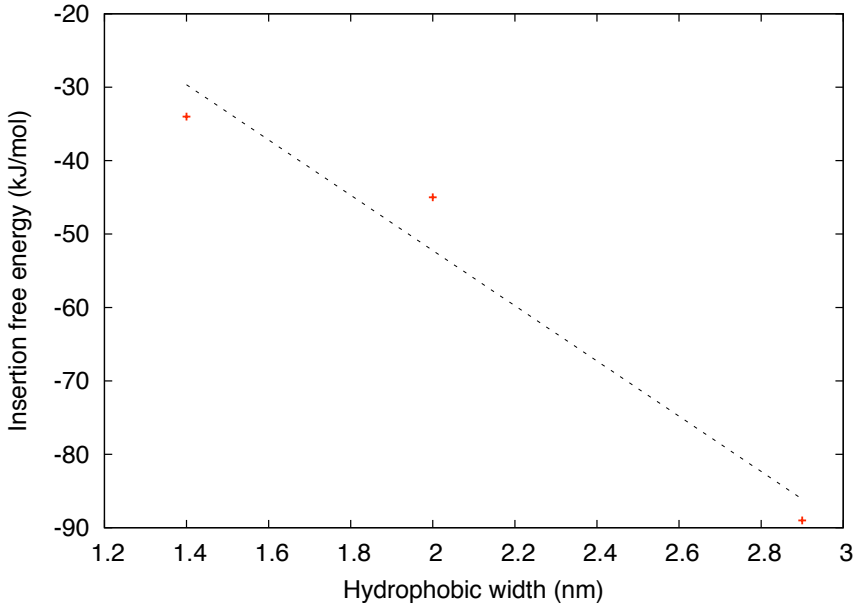


Figure 3.16: Plot of ΔG_{pmf} with (unperturbed) hydrophobic bilayer width. A dashed straight line of best fit is shown in black, with a correlation coefficient of -0.98.

other words fewer of the windows start with a configuration where both ends of the helix traverse the headgroup region. Where one end of the helix starts in the hydrophobic core of the membrane, the end is rapidly expelled, often resulting in the helix adopting an IF state. This explains the relative scarcity of IF states in the thinner bilayers compared with the DPPC simulations (see also Figure 3.17).

When the N-terminal end of the helix is facing the centre of the bilayer (i.e. when the distance from bilayer centre is negative) the helix is not significantly tilted in the bilayer, as can be seen in Figure 3.17. However, as the simulations start with the N-terminal end out of the bilayer (i.e. when the distance from the bilayer centre is positive, ~0-2 nm), the charged residues in positions 2, 4, 5, 6 and 7 seek to interact with the bilayer headgroups rather than water. This causes the helix to tilt significantly, resulting in a larger tilt angles as the distance between the helix and bilayer centres of mass becomes more positive (Figure 3.17).

3.3.3 Initial Orientation

To further test the sampling in the PMF simulations I calculated the PMF of M2-39 across a lipid bilayer using a starting orientation with the helix lying parallel with the membrane (M2-39_{90}), rather than perpendicular to the membrane as with previous simulations. The

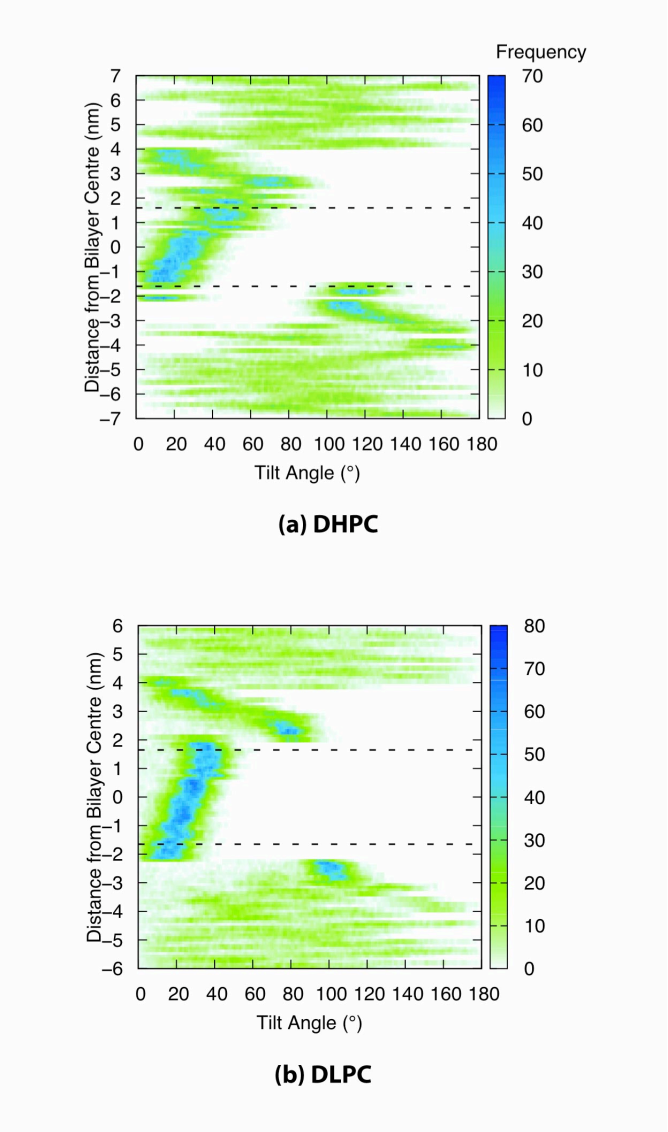


Figure 3.17: The distribution of helix tilt angles from the M2-39 simulations with (a) DHPC and (b) DLPC bilayers. Zero degrees corresponds to the helix being perpendicular to the plane of the membrane. The dashed black lines show the approximate edges of the bilayers.

resulting PMF is shown in Figure 3.18, together with the profile of M2-39 with the helix starting perpendicular to the bilayer (M2-39_0). This shows that the PMFs are similar when the helix is in water and in an IF position - $\Delta\mathcal{W}_{WI}$ is the same for both starting positions. However, there is a clear difference in the centre of the bilayer, with the M2-39_{90} profile showing a barrier but the M2-39_0 profile showing a deep well. In the former case, the helix gets kinetically trapped in a position where most of the residues remain in the hydrophobic core of the membrane. The helix adopts some energetically costly bending into a u-shape to allow the polar and charged residues to interact with the lipid headgroups (see Figure 3.19), but there is too high a barrier for the helix to flip into a fully transmembrane position (see Figure 3.20). Hence the helix gets stuck in a relatively unfavourable position. The suggested energy landscape is sketched in Figure 3.21. Preventing the u-shaped state shown here is one possible explanation for the evolution of the translocon's 'perpendicular' mechanism of inserting TM helices into the bilayer. Membrane proteins typically require most helices to traverse the membrane completely, and inserting the helices perpendicular to the membrane, rather than via an intermediate IF state, prevents kinetic trapping in such a u-shape.

The M2-39_{90} profile is not continuous; at the very centre of the bilayer no PMF values have been calculated. This is due to poor sampling in this region. Histograms showing the frequency of helix centre of mass positions with respect to the bilayer centre are shown in Figure 3.22. This shows that the helix centre of mass is much more reluctant to sample the bilayer centre in the M2-39_{90} simulations than in the M2-39_0 simulations, although the spacing of the centre of the restraint was the same for both calculations. This feature is a result of the whole helix being trapped in a u shape in the centre of the bilayer in the M2-39_{90} simulations, and being unable to flip to adopt a TM position.

The fact that this kinetic trapping can occur, biased by the starting position of the helix, suggests that these PMFs need to be understood as showing one particular path through the energy landscape of the system. The behaviour of the helix in each window needs to be assessed to ensure it is sampling a region of phase space which will be near an energy minimum.

3.3.4 Folding State

Secondary structure cannot be formed in the course of a Bond CG simulation as the spherically symmetric CG particles cannot reproduce the directionality of hydrogen bonds in α-helices. Instead secondary structure is modelled using a harmonic restraint applied accord-

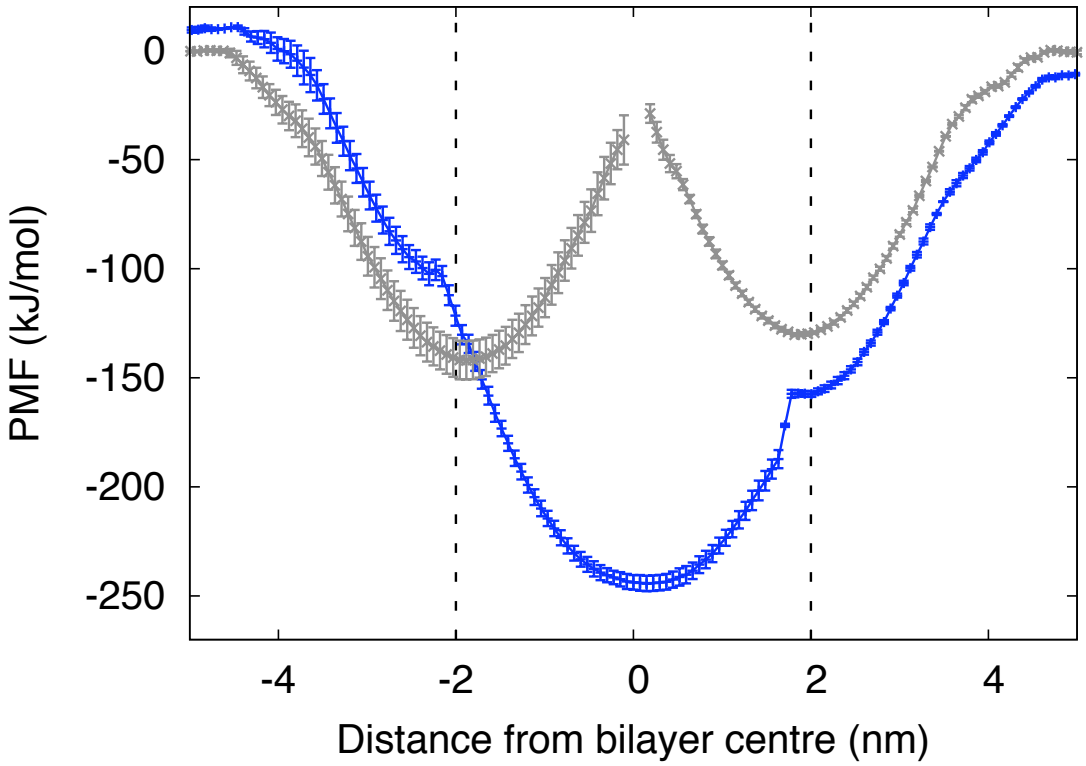


Figure 3.18: PMF profiles for the transfer of M2-39 across a DPPC bilayer with different starting orientations with respect to the bilayer. Blue: helix starts perpendicular to the bilayer. Grey: helix starts parallel with the bilayer.

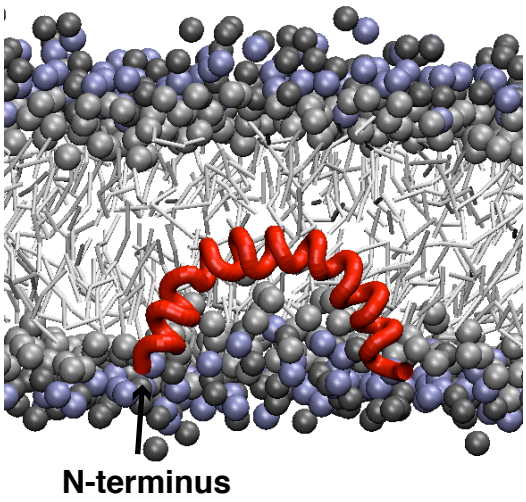


Figure 3.19: Snapshot of M2-39_{90} in a bilayer. The helix adopts an upturned u-shape to allow the polar ends of the helix to interact with the lipid headgroups.

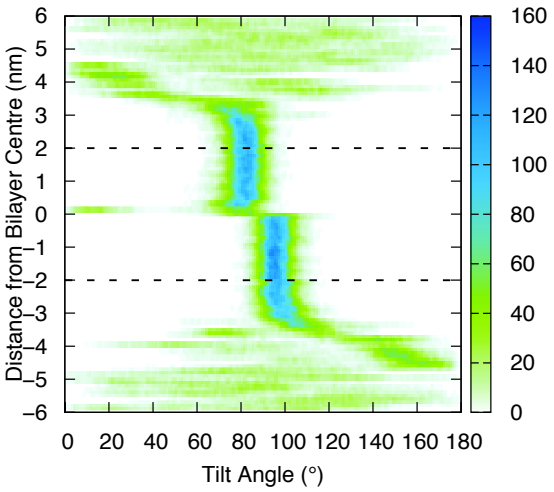


Figure 3.20: Tilt angle of M2-39$_{90}$ as a function of distance from the bilayer.

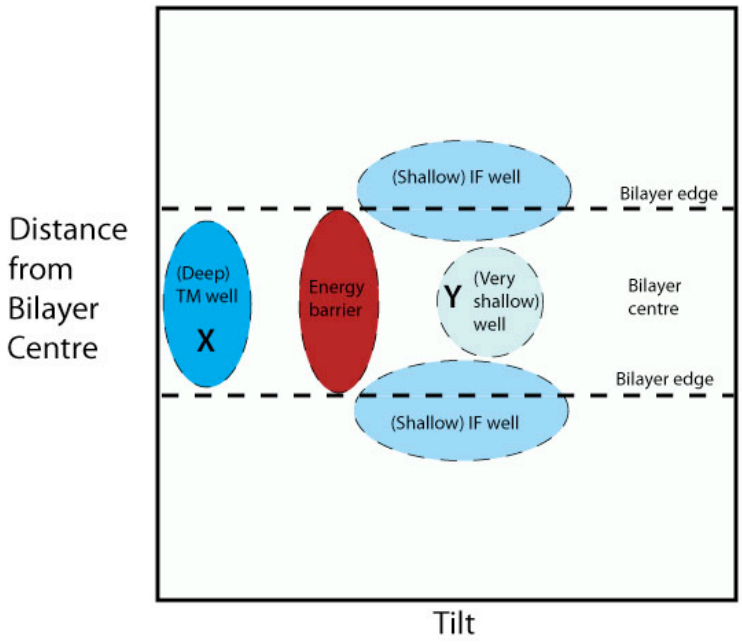
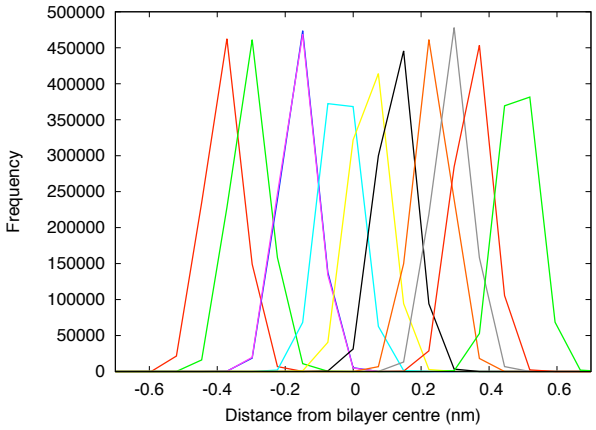
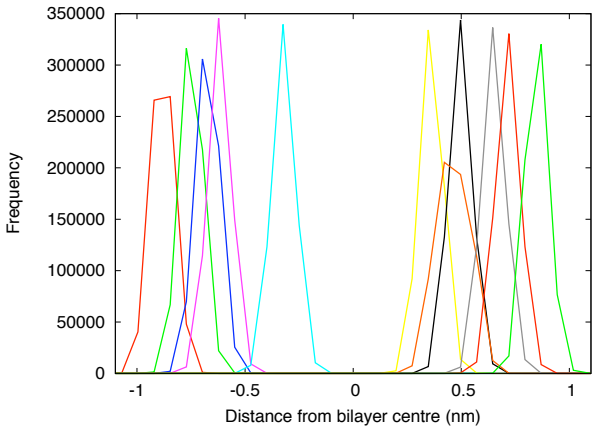


Figure 3.21: Schematic of suggested energy landscape for M2-39 as a function of helix distance from bilayer and helix tilt with respect to bilayer normal. A helix starting at an angle of 0° (position X) will stay in the TM energy well, while a helix starting at an angle of 90° (position Y) will stay kinetically trapped in the very shallow well corresponding to the u-shape in the centre of the bilayer.



(a) M2-39_0



(b) M2-39_{90}

Figure 3.22: Histograms showing the frequency of displacement of helix centre of mass near the bilayer centre for the M2-39_0 (top) and M2-39_{90} (bottom) simulations.

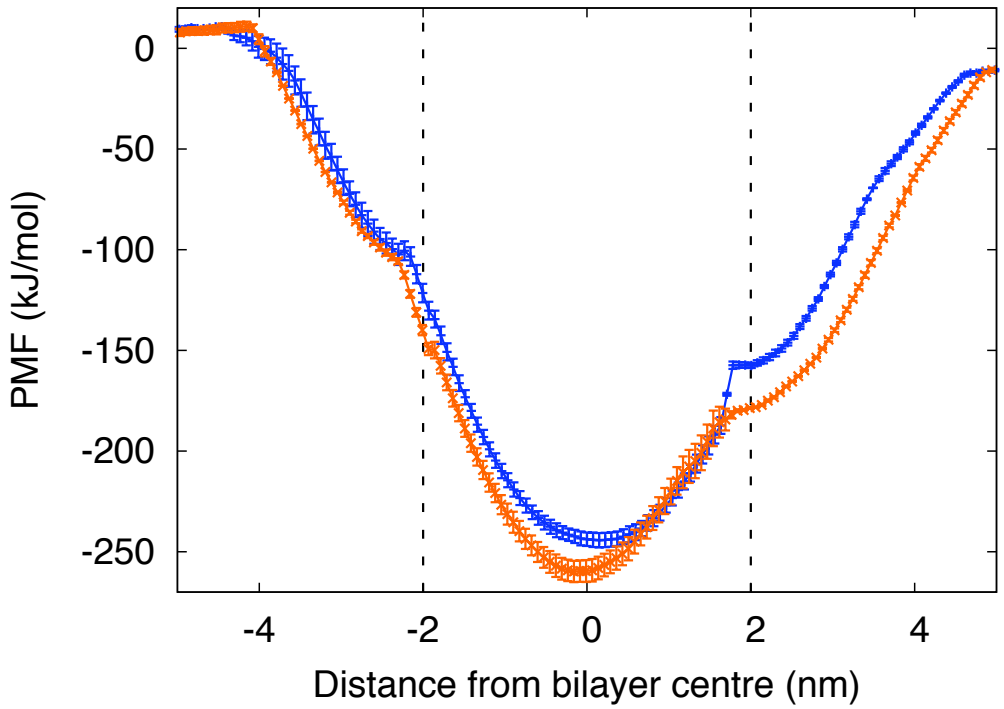


Figure 3.23: PMF profiles for the transfer of M2-39 across a DPPC bilayer with different secondary structure restraints. Blue: helical restraints applied to full sequence. Orange: helical restraints applied to M2-19 only.

ing to a distance cut-off (0.7 nm) in the initial structure. The Bond CG model is therefore unable to model transitions between different folding states during simulations. However, as the folding state of the investigated sequence is unclear during the translocon mediated insertion experiments, it was important to assess the effect of helix folding state on the PMF. I therefore calculated the PMF of M2-39, maintaining these helical restraints only on the sequence corresponding to M2-19. The rest of the M2-39 sequence was free to move flexibly, similar to a coil region (see Figure 3.24). This structure is termed M2-39$_{partial}$. The resulting PMF is shown in Figure 3.23. The profiles look very similar, with the $\Delta\mathcal{W}_{wc}$ and $\Delta\mathcal{W}_{ic}$ almost unaffected by the different secondary structure restraints. The values of ΔG_{pmf} are also not significantly affected, at -89 kJ/mol for the ideal helix and -82 kJ/mol for the partial helix. This shows that the flexibility of the edges of the sequence do not affect the PMF significantly. However, it does not include the effect of total unfolding of the helix in bulk water, which we might expect to make ΔG_{pmf} more positive.

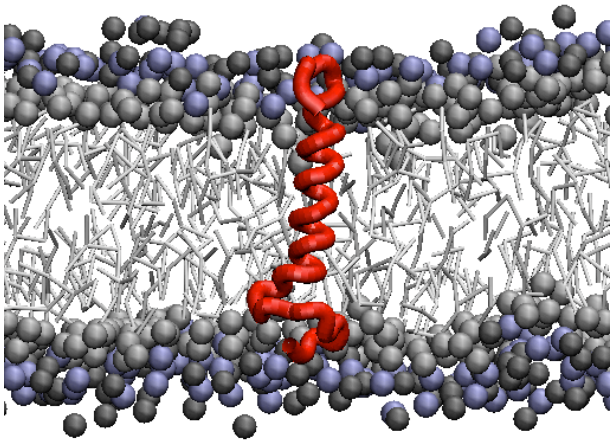


Figure 3.24: Snapshot of M2-$39_{partial}$ in a bilayer. The unstructured ends of the helix can adopt a flexible position with respect to the bilayer.

Table 3.5: M8 Sequence

Helix	Sequence
M8	KGNSTHSRNNSYAVIITSTSSYYVFYIYVGVADTLLAMG

3.3.5 Helix Hydrophobicity

Although hydrophobicity is known to be a vital element in membrane protein folding, some membrane proteins contain helices which are only marginally hydrophobic [200]. Enquist and colleagues identified the 8th transmembrane helix from CFTR (M8) as one such helix which does not insert independently through the translocon [191]. M8 only shows an insertion efficiency of greater than 50% when an extra hydrophobic region at the C-terminal end is included in the investigated sequence. I therefore calculated the PMF of the 39 residue sequence surrounding M8 (M8-39, see Table 3.5). The resulting profile is shown in orange in Figure 3.25. This showed an unphysical difference of ~100 kJ/mol in the energy levels of the IF states above and below the bilayer, and no energy well in the centre of the bilayer. To obtain an alternative estimate for the energies of the IF states I therefore calculated the PMF profile of M8-39, starting the helix parallel to the bilayer (M8-39_{90}). The resulting profile is shown in turquoise in Figure 3.25. This showed similar energy levels for the IF states above and below the bilayer, as might be expected; this suggests a problem with the original M8-39_0 calculation. It is interesting to note that snapshots from both M8-39_0 and M8-39_{90} show similar IF and TM states (Figure 3.27).

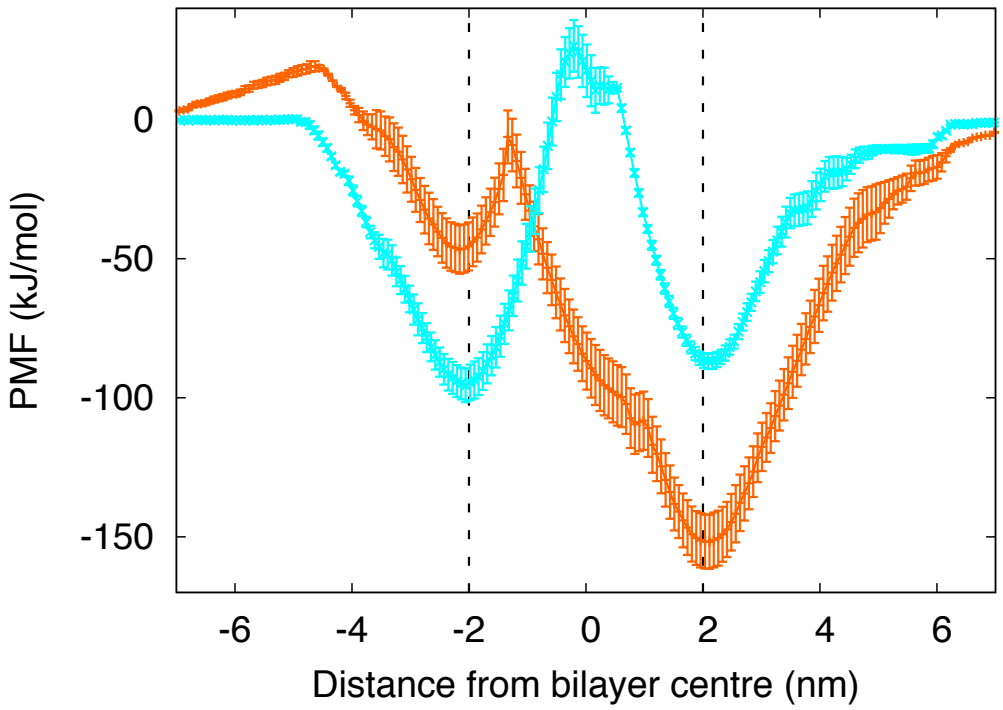
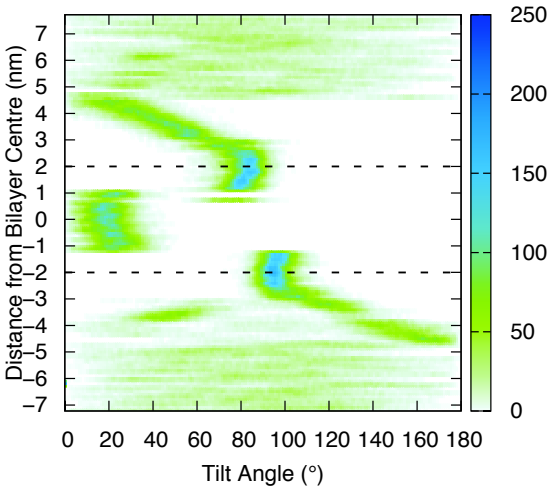
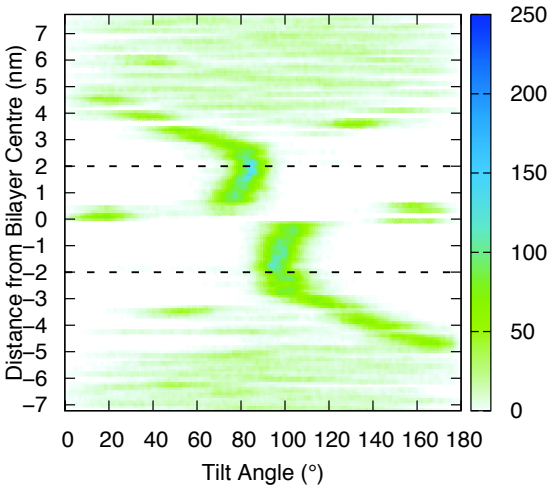


Figure 3.25: PMF profiles for the transfer of M8 across a DPPC bilayer. Orange: helix starts perpendicular to the bilayer (M8-39_0). Turquoise: helix starts parallel with the bilayer (M8-39_{90}).



(a) M8-39$_0$



(b) M8-39$_{90}$

Figure 3.26: Frequency of helix tilts as a function of reaction coordinate for (a) M8$_0$ and (b) M8$_{90}$.

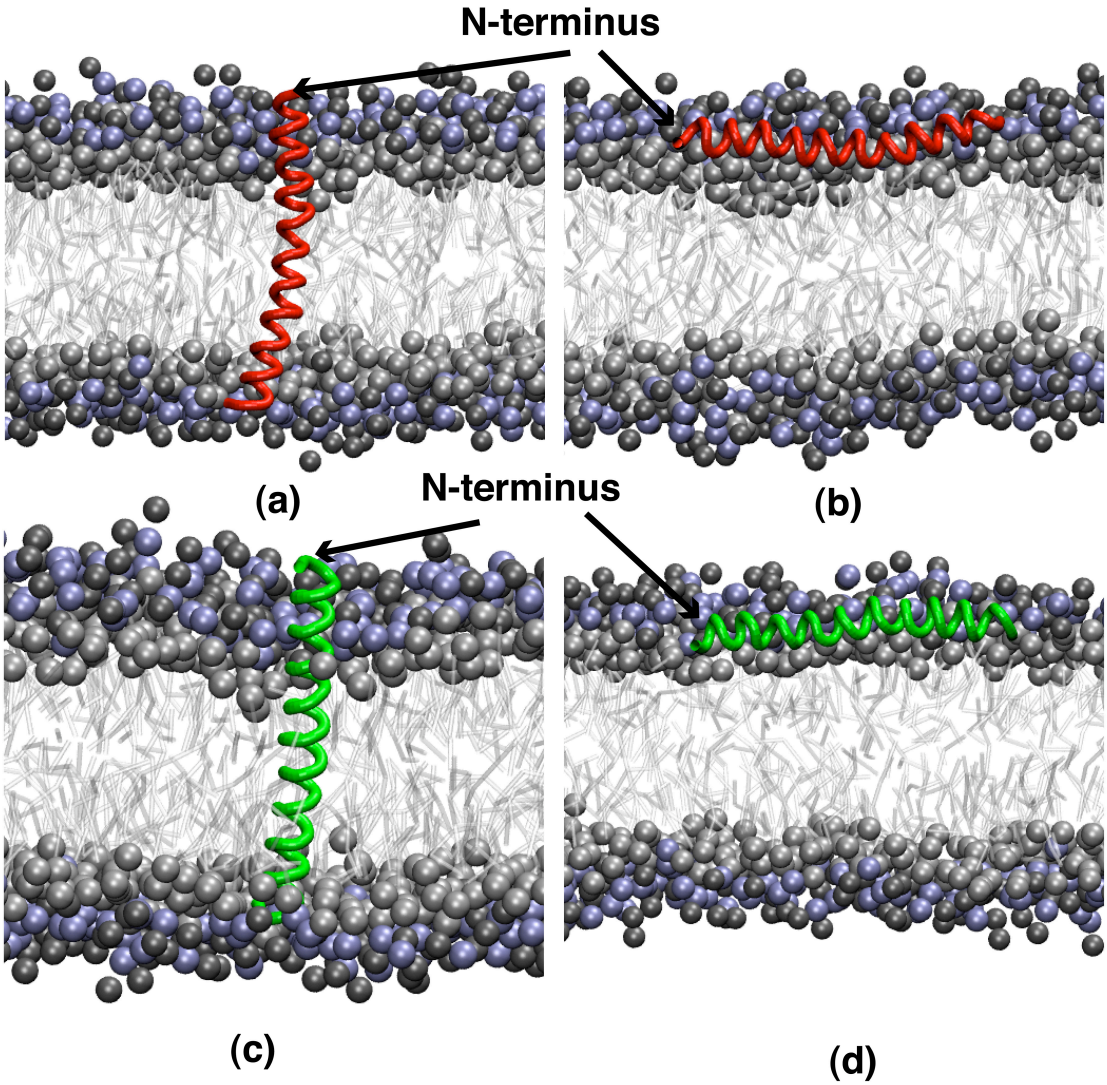


Figure 3.27: Snapshots showing the TM and IF states of the M8-39_0 (red helix) and M8-39_{90} (green helix) simulations. Pictures taken from simulations with a restraint applied to (a) 0.2 nm, (b) 2 nm, (c) 0.1 nm, (d) 2 nm separation between the helix and bilayer centres of mass.

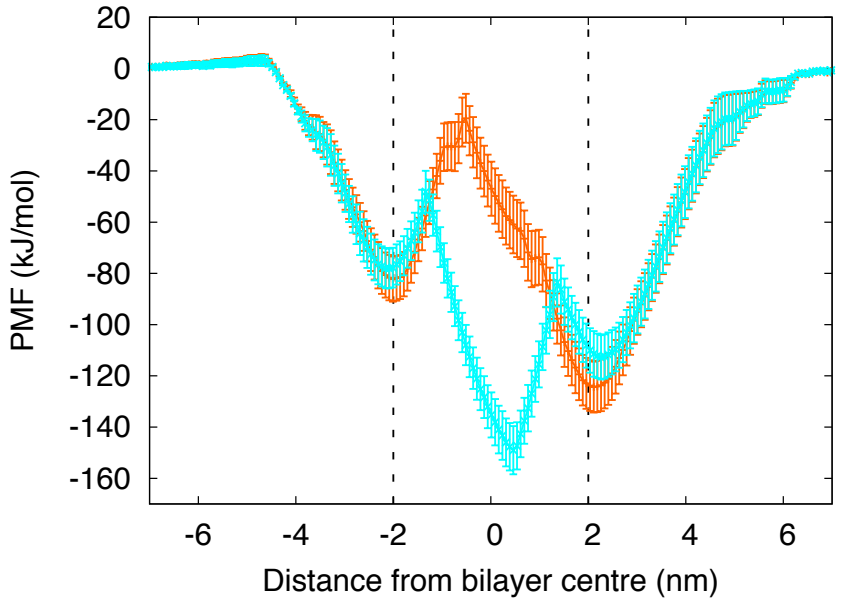


Figure 3.28: PMF profiles for the transfer of M8 across a DPPC bilayer. Orange: displacements from 26 windows on the left have been replaced. Turquiose: displacements from 26 windows on the right have been replaced.

The large difference in IF energies shown in the PMF profile for M8-39_0 would seem to suggest that the IF state is adopted more frequently on one side of the bilayer. However, investigation of the helix tilt angle as a function of reaction coordinate (Figure 3.26) shows that the helix behaves similarly on either side of the bilayer. Attempts to replace the distribution of helix displacements of selected windows with their 'equivalent' on the other side of the bilayer provided profiles which did not seem physically reasonable (Figure 3.28), as these profiles were also very asymmetric, with different energies in the different IF states and no discernible TM well.

Taking the average of the original PMF on either side of the bilayer (Figure 3.29) produces a profile which seems physically reasonable for a marginally hydrophobic helix. This profile shows the IF states are a little more stable than the TM state. However, the justification for such symmetrising of a profile of an asymmetric helix is uncertain. The PMF profiles previously presented have been derived from simulations where the N-terminus of the helix points in a consistent direction in the starting configuration - for a particular profile, the negative values of the reaction coordinate refer to simulations which start with the N-terminus pointing towards (or away from) the bilayer. It has been shown that this starting structure can have a fundamental effect on the subsequent simulation and the cal-

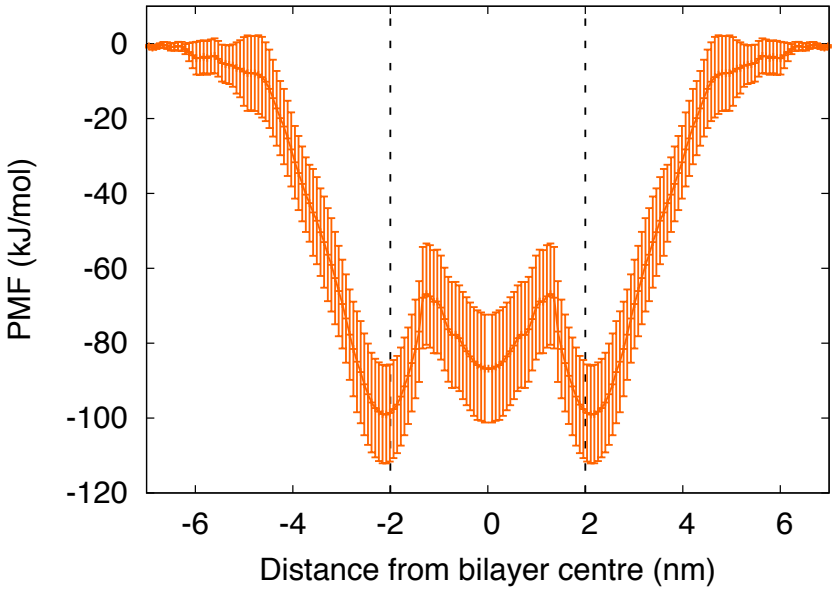


Figure 3.29: PMF profiles for the transfer of M8 across a DPPC bilayer. The profile has been symmetrised across the bilayer, with error bars produced from the rms of the difference between first and second halves of the simulation and PMFs from either side of the bilayer.

culation of the PMF (see for example Figure 3.12). The PMF generated therefore describes a particular entry pathway into (or exit pathway out of) the bilayer. This profile is therefore dependent on the sequence of the helix, where an asymmetric sequence can lead to asymmetric behaviour, which in turns yields a profile with some level of asymmetry (see for example Figure 3.3). However, to symmetrise a profile means that simulations where the helix is restrained on a different side of the bilayer are treated as equivalent, regardless of the initial angle of the helix. In other words, the reaction coordinate becomes simply the separation between the bilayer and helix centres of mass. Defining the reaction coordinate in this way requires confidence that the bias from the starting structure has been removed. This can only be achieved with a large number of simulations using a variety of starting positions. Using two simulations for each value of the reaction coordinate, with the helix starting angle differing by 180° in each simulation, is unlikely to be sufficient to remove this bias. Therefore symmetrising the M8-39 profile may be able to provide a useful estimate of the PMF with the alternative definition of the reaction coordinate, but the errors involved are likely to be too large to give an accurate estimate for ΔG_{pmf}.

3.4 Discussion

3.4.1 Sampling

These PMFs are an accurate description of the free energy landscape of the system along the chosen reaction coordinate only if all degrees of freedom have been sufficiently sampled. For the PMF calculations in this chapter, this means that the helix must be able to adopt its most energetically favourable orientation with respect to the bilayer in all simulations. It is reasonable that this would mean the helix must tumble freely when in bulk water, sampling all tilts equally; that transmembrane helices would adopt a fully transmembrane position in the centre of the membrane; and they would adopt a more complex position, possibly lying along the headgroup region of the bilayer, when placed near the interface between the bilayer and water. It is apparent in all calculations that the helix does indeed tumble freely in water (see for example Figure 3.26). However, it is apparent that the choice of starting structure limits the position which the helix may adopt when it is in the bilayer. For instance, M2-39_0 adopts a transmembrane position in most simulations, indicating a deep energy minimum at the centre of the bilayer (Figure 3.18); but in the M2-39_{90} simulations, it gets trapped in the bilayer centre, and so the profile shows an energy barrier at the centre of the bilayer. The two different starting positions therefore ensure that the system samples different regions of the energy landscape, so the resulting profiles show different 1-dimensional paths through the multidimensional free energy landscape for the system.

This can be understood by viewing the PMF as the projection of the free energy landscape onto a chosen reaction coordinate. The reaction coordinate used for these calculations involved the helix moving across the bilayer starting each window with its N-terminal end pointing in a consistent direction. This means that the profiles generated will not in general be symmetric; with the helix pointing in a consistent direction, a helix with an asymmetric sequence may behave differently depending on whether for example a more charged end is closer to or further away from the bilayer. Such asymmetric behaviour can be seen in Figure 3.8.

In some profiles the energy of the helix in bulk water on different sides of the bilayer is very different (see for example Figure 3.30). This is probably a result of kinetic trapping combined with the requirement from WHAM of the profile to be continuous. If the helix adopts a slightly different state on one side of the bilayer compared with the equivalent simulation on the other side, it is likely that the helix has become kinetically trapped in one

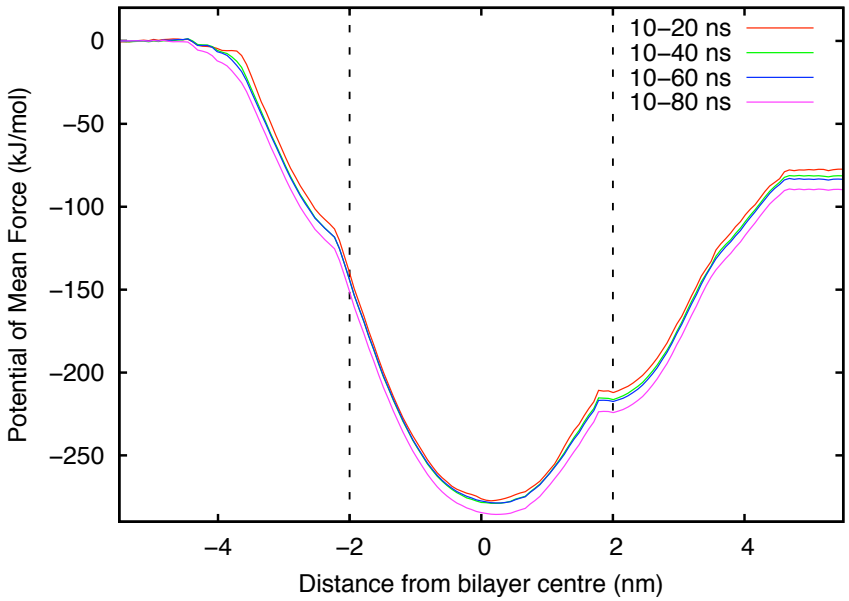


Figure 3.30: PMF profiles for the transfer of M2-39 across a DPPC bilayer calculated for gradually increasing times of the simulation with the end simulations not clamped. According to these profiles the helices in water are at different energies whether above or below the bilayer, which is clearly unphysical. This is why other profiles are 'clamped'.

of the simulations and is unable to find its most stable state at that point on the reaction coordinate. This may cause the energy of two equivalent points on the reaction coordinate to look rather different. Sometimes this kinetic trapping cannot be easily identified and emerges as a difference in energy in two equivalent states on the PMF, such as the helix in bulk water.

Kinetic trapping dependent on the starting structure can also create PMF profiles which are not physically reasonable, as in Figure 3.10. In such instances an understanding of the expected behaviour of the system can allow us to use equivalent data which are likely to represent a lower energy state. For the M2-29 PMF calculation, the behaviour of the helix was different when it was restrained along the reaction coordinate from -1.5 nm to 0 nm from the bilayer compared with 0 nm to 1.5 nm. This seemed to be caused by snorkelling of the charged Arg^{24} residue drawing the helix into IF or TM states depending on the starting structure. As M2 is a transmembrane helix and the M2-39 PMF profile indicated a large TM well, it seemed reasonable to replace the displacement values from the 'biased' IF windows with the inverted displacement values from the equivalent windows on the other side of the bilayer. This did indeed provide a PMF profile which seemed physically reasonable, with IF

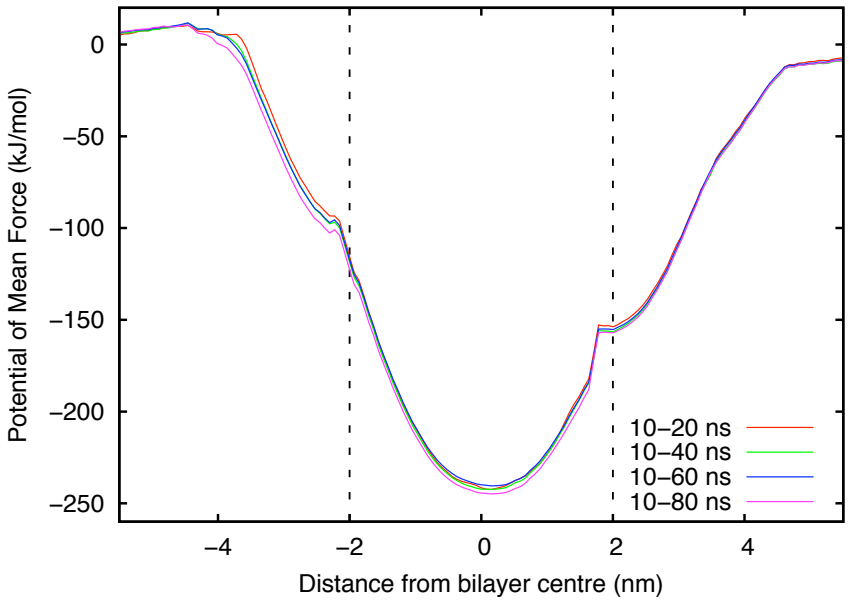


Figure 3.31: PMF profiles for the transfer of M2-39 across a DPPC bilayer calculated for gradually increasing times of the simulation, with the end simulations 'clamped' at 0 kJ/mol. Profiles do not change significantly whether calculated from 10-20 ns or 10-80 ns, suggesting that sampling over tens of nanoseconds is adequate to calculate a PMF profile along this reaction coordinate.

wells less deep than the TM well and ΔG_{pmf} values comparable with those obtained from calculations of M2-39.

Given that this kinetic trapping can bias the PMF, how can we be confident that the profiles in this chapter are indeed valid? Calculations of the profiles over increasing time periods of the simulation windows show that the profiles do not change significantly whether the PMF is calculated over a 10 ns period or a 70 ns period (see for example Figure 3.31). The differences between the PMFs over the first or second half of equilibrated simulation time are small. Analysis of tilt angles shows that the helices tumble freely in water over a $\sim$5 ns timescale. Results from the PMFs are also consistent with results attempted on model peptides and a variety of different hydrophobicity scales (see below).

Ideally we would like to be able to calculate the full free energy landscape of a given system. This would involve either prohibitively long simulations, or a PMF calculation restraining a large number of degrees of freedom. Alternatively a sufficiently large number of simulations with different starting positions may be able to overcome kinetic trapping in most windows, so sample all degrees of freedom adequately. However, all of these approaches are extremely computationally expensive. Given limited computational resources,

we can interpret 1-dimensional PMFs in terms of one particular path which the simulations trace across the energy landscape. It may be that the path taken in one dimension will not always sample the energy minimum at that point along the reaction coordinate.

It would be possible to prevent concerns over sampling of a particular degree of freedom by fixing that degree of freedom in all simulations. For instance, the helix could be restrained in a perpendicular position with respect to the bilayer such that it could not rotate; however, this would mean the helix would be unable to adopt an IF position and so we would not be able to obtain accurate ΔG_{pmf} values to compare with experiment.

Treatment of errors

Error bars shown in the profiles are the difference between the PMF calculated from the first and second halves of equilibrated simulation time. Errors on ΔG_{pmf} have not been estimated. A rigorous estimate of the statistical error on ΔG_{pmf} would involve calculating the correlation time τ for each profile, by calculating values of ΔG_{pmf} over different periods of equilibrated simulation time. The shortest time at which values of ΔG_{pmf} remain constant is τ. Simulation data can then be divided into independent sets each of length τ, and a series of independent values of ΔG_{pmf} can then be calculated from each of these data sets. The statistical error on ΔG_{pmf} is then the standard deviation of the calculated values of ΔG_{pmf}, divided by the square root of the number of values of ΔG_{pmf}. These would be time-intensive calculations, and would not include systematic sources of error e.g. errors in the parameterisation of the forcefield or inadequate sampling due to kinetic trapping in a simulation.

3.4.2 Comparison of Results

Comparison with Hydrophobicity Scales via an Additive Model

To ensure that the insertion free energies obtained from the CG PMF calculations are reasonable, they were compared with simple estimates using hydrophobicities derived from a variety of free energy scales. The model used is to sum all of the individual ΔG_i values for every residue in the bilayer core when the helix is in the TM state. ΔG_i is the free energy of insertion of the ith residue of the helix as estimated from each hydrophobicity scale. The residues in the bilayer core - i.e. those interacting with the lipid tails - when the helix is in the TM state were determined by inspection.

The available hydrophobicity scales report two different quantities. The experimental

Table 3.6: Estimates of ΔG_{ins} obtained from summing the per residue ΔGs from different hydrophobicity scales. ΔG_{R-W} : Radzicka-Wolfenden scale from partitioning amino acid side chains from water into cyclohexane [164]. ΔG_{TR}: Obtained from translocon-mediated insertion experiments [35]. ΔG_{EL}: obtained from continuum electrostatics calculations [181]. ΔG_{W-W}: Wimley-White scale obtained from the partitioning of whole residues from water to octanol [37]. ΔG_{K-D}: Kyte-Doolittle scale from partitioning from water to vapour and probability of a residue to be found in the hydrophobic core of a test set of proteins [185].

Helix	ΔG_{R-W}	ΔG_{TR}	ΔG_{EL}	ΔG_{W-W}	ΔG_{K-D}	Average	ΔG_{pmf}
M2-19	-162	-2	-153	-36	-170	-105	+24
M2-29	-143	17	-140	-35	-147	-90	-49
M2-39	-221	11	-200	-52	-211	-135	-89
M2-39 (DLPC)	-135	17	-133	-37	-139	-85	-45
M2-39 (DHPC)	-110	-3	-113	-30	-134	-78	-34
M8-39	-95	17	-133	-25	-120	-71	+45

Table 3.7: Estimates of $\Delta \mathcal{W}_{wc}$ obtained from summing the per residue $\Delta \mathcal{W}_{wc}$s from different hydrophobicity scales. CG-B: Bond CG model [1]. CG-M: MARTINI CG model Monticelli:2008. AT: Obtained from atomistic PMF calculations [201].

Helix	CG-B	CG-M	AT	Average	$\Delta \mathcal{W}_{wc}$
M2-19	-192	-209	-117	-173	-68
M2-29	-175	-210	-110	-165	-166
M2-39	-263	-295	-183	-247	-244
M2-39 (DLPC)	-156	-210	-102	-156	-212
M2-39 (DHPC)	-141	-166	-78	-128	-176
M8-39	-103	-106	-50	-86	-85

and theoretical scales report a ΔG_{wc}, a free energy change on insertion of the residue from water into some hydrophobic environment. The computational scales typically report a PMF depth $\Delta \mathcal{W}^i_{wc}$, a difference in the PMF of the ith residue in water and in the centre of a bilayer. To make sure similar quantities are compared, results from the additive model using free energy scales reporting ΔG_{wc} are compared with ΔG_{pmf} values from simulations (Table 3.6). Results from the additive model using $\Delta \mathcal{W}^i_{wc}$ values are compared with the PMF depth $\Delta \mathcal{W}_{wc}$ from simulations (Table 3.7).

This model is obviously extremely simplified. Free energy is not simply additive, and the model neglects important interactions in the headgroup region as well as the positional dependence of ΔG_i values [36]. Nevertheless, as the free energies of insertion measured computationally and from translocon mediated insertion experiments typically differ by close to an order of magnitude, such a simple model may be able to provide a valuable

check that computational methods provide reasonable results.

Estimates of $\Delta\mathcal{W}_{wc}$ using the additive model ($\Delta\mathcal{W}_{wc}^{add}$) and various computationally derived scales compare well with the values obtained from the PMF calculations. The average estimate taken across all three scales compares very well with the PMF results for M2-29 (-165 for additive model, -166 from PMF), M2-39 (-247 / -244), and M8-39 (-86 / -85). Values for M2-39 in DHPC and DLPC bilayers are also comparable, given the simplicity of the additive model. It is only for M2-19 where the additive model gives a significantly different result from the PMF; the model suggests a $\Delta\mathcal{W}_{wc}^{add}$ of -173, where the value from simulations is -68. This is most likely because the additive model doesn't take into account the significant energy cost of bilayer distortion to match the hydrophobic thickness of a helix.

The average estimates of $\Delta\mathcal{W}_{wc}^{add}$ are significantly larger for the two CG scales (-177 and -199 kJ/mol) than for the atomistic scale (-106 kJ/mol). This may well be due to the treatment of electrostatics: in the CG models, electrostatic interactions are treated with a cutoff of 1.2 nm, whereas in the atomistic model electrostatics is treated to a longer range using PME. This means that a charged residue in the centre of the bilayer in the CG model would not experience the electrostatic interactions with lipid headgroups, so the energetic cost of placing the residue in the bilayer centre is likely to be under-estimated. This may explain why the CG scales produce higher $\Delta\mathcal{W}_{wc}^{add}$ estimates than the atomistic scale.

The agreement between ΔG_{pmf} values from PMF calculations and estimates of ΔG from the additive model (ΔG^{add}) using hydrophobicity scales from Table 3.6 is slightly less good than the estimates using computational scales. The dependence of ΔG_{pmf} on bilayer width generally is reflected in ΔG^{add}, and the dependence on helix length is shown in results for M2-29 and M2-39. However, the results for the additive model for M8-39 and M2-19 are not similar to ΔG_{pmf} values from PMF calculations. This is partly due to the additive model ignoring the energetic cost of M2-19 distorting the bilayer, but also due to the prevalence of the IF state in the M2-19 and M8-39 simulations. Both show large energy wells in the IF state, which are taken into account in the ΔG_{pmf} values, which measure 'inserted' against 'not-inserted' probabilities for the helix. However, for the additive model the inputs are individual ΔG values taking residues from water into a hydrophobic environment. These do not take into account the possibility of an IF state, so the ΔG_{ins} values from the additive model are likely to be too negative for helices with a significant IF energy well.

The Radzicka-Wolfenden, Choe and Kyte-Doolittle scales provide estimates of ΔG_{ins}^{add} which are similar. All estimates using these scales are too large, probably as the scales do

not take into account the IF state. The estimates derived from the Wimley-White scale are very much smaller than those from the other scales, either due to the use of octanol as a mimic for the membrane environment - it has been suggested that octanol is a better mimic of the interfacial region [182] - or because they included the cost of putting the peptide bond into the membrane, which is not included in the CG simulations owing to the inability to model folding transitions. The translocon scale provides estimates of ΔG which do not match with ΔG_{pmf} values from PMFs. In fact the estimates are mostly positive and do not show the same length and bilayer width dependences as the other scales, and show no correlation with estimates from the other scales. This may be because the individual ΔG_i values are so small that the errors from the positional dependence of the residues are larger than the ΔG^{add} helix length or bilayer width dependence. It may support the view that the translocon hydrophobicity scale describes a quantity which is dependent on the path the residue takes from the ribosome into the bilayer, and is not necessarily a good description of the absolute free energy cost of taking the residue from water and putting it in a hydrophobic environment.

A major flaw in the simple additive model is that it neglects interactions between the helix and the bilayer headgroups. This can be estimated using three of the hydrophobicity scales. The translocon-mediated experiments assessed the position dependence of the amino acid ΔGs, including near the headgroup region; the atomistic PMF calculations included the free energy of the residue in the headgroup region as well as the centre of the bilayer; and Wimley and White used POPC vesicles, calculating the partitioning of whole residues from water into the POPC headgroup region [202]. Results from including these 'interfacial' hydrophobicity scales in the calculations are shown in Table 3.8. It might be expected that including the interactions of the more polar residues on the edges of the helix sequences with the bilayer headgroups would result in more favourable free energies of insertion, but in fact only the atomistic free energies of insertion are consistently reduced relative to the ΔG^{add} values calculated using just the interactions with lipid tails. Estimates from the Wimley-White and translocon derived scales increase when the interfacial interactions are taken into account. This may reflect the diversity of chemical behaviours of different lipid headgroups and lipid mixtures, and the difficulty in reproducing this behaviour. Overall it seems that the additive model is not improved by attempting to include the effect of interfacial interactions.

Table 3.8: Estimated insertion free energies from additive model using IF free energies. Estimates of ΔG_{ins} obtained from summing the per residue ΔGs from different hydrophobicity scales, including individual ΔGs for taking IF residues from water into a region with similar hydrophobicity to a bilayer headgroup region. ΔG^{TR}: Obtained from translocon-mediated insertion experiments [35]. ΔG^{W-W}: Wimley-White scale obtained from the partitioning of whole residues from water to octanol [37]. $\Delta \mathcal{W}^{at}$: Obtained from atomistic PMF calculations [201].

Helix	$\Delta \mathcal{W}^{at}$	ΔG^{tr}	ΔG^{ww}	$\Delta \mathcal{W}_{wc}$	ΔG_{pmf}
M2-29	-219	15	-34	-68	-49
M2-39	-281	22	-18	-166	-89
M2-39 (DHPC)	-294	6	-6	-244	-34
M2-39 (DLPC)	-277	22	-16	-212	-45
M8-39	-157	28	-12	-85	+45

Comparison with other PMFs

The PMF profile for transfer of WALP23 across a DPPC bilayer has previously been calculated using the Bond CG model and umbrella sampling [1]. WALP23 is a hydrophobic model helix consisting of alternating Ala and Leu residues flanked by Trp. The profile generated suggested a value of $\sim$-194 kJ/mol for $\Delta \mathcal{W}_{wc}$, which compares with values of -166 and -244 kJ/mol for M2-29 and M2-39 respectively. As WALP23 is slightly shorter but significantly more hydrophobic than M2-29 it is reasonable that it should have a slightly larger $\Delta \mathcal{W}_{wc}$. The value of $\Delta \mathcal{W}_{ic}$ is $\sim$-32 kJ/mol for WALP23, comparing with -47 kJ/mol for M2-29.

Gkeka and Sarkisov used the MARTINI model to calculate PMF profiles for a variety of model peptides: pore forming, amphipathic, fusion and transmembrane peptides [189]. Different profiles were produced depending on the distribution and hydrophobicity of the sequences. Barrel-stove forming peptides showed wells of equal energy in IF and TM states; amphipathic peptides showed energy wells in the IF states but a massive energy barrier in the TM state; a fusion peptide showed wells in IF states and a small barrier in the TM state; and the transmembrane peptides showed no significant well in the IF state but a deep well in the TM state. The ΔG_{wc} for the transmembrane peptides were $\sim$-243 kJ/mol and $\sim$-297 kJ/mol for the peptides N-pHLIP and TMX-1 respectively, which compare with $\Delta \mathcal{W}_{wc}$ of -244 kJ/mol for M2-39.

It is therefore apparent that both the shape of the PMF profiles of M2-19, M2-29 and M2-39, and their well depths $\Delta \mathcal{W}_{wi}$ and $\Delta \mathcal{W}_{ic}$ are of comparable magnitudes to the similar calculations that have previously been reported in the academic literature.

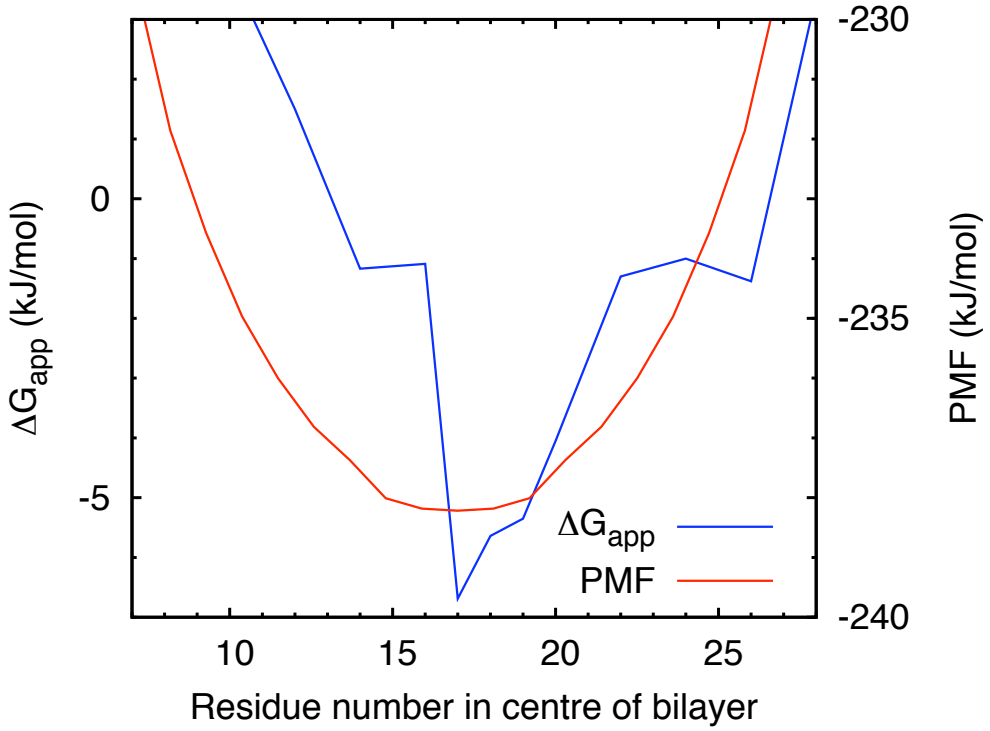


Figure 3.32: M2-39 PMF (red) and ΔG_{app} values from 19 residue windows from M2, measured by translocon-mediated insertion experiments, are shown as a function of the central residue number of the helix investigated (blue). Both techniques show a similar qualitative dependence of helix stability on sequence.

Comparison with Experiment

A qualitative comparison of PMFs with the ΔG_{app} values measured in translocon-mediated experiments shows notable similarities. The PMF for M2-39 shows a minimum when the centre of mass of the helix is ∼0.15 nm away from the centre of the bilayer (towards the N-terminus of the sequence). The ΔG_{app} values measured by Enquist et al were at a minimum for the M2-19(-1) sequence. The centre of this sequence corresponds to Cys^{18}, which is ∼0.3 nm closer to the N terminus of the sequence than its central residue Leu^{20}. The most stable state for M2-39 as measured by PMF simulations therefore corresponds to the highest-inserting sequence as measured by experiment in the centre of the membrane.

A further comparison can be made by plotting the M2-39 PMF as a function of the residue number at the centre of the bilayer. This can be compared with the ΔG_{app} values for the H-segments with this residue at the centre of the sequence. This is by no means an exact comparison, but both values correspond to a measure of how stable the helix is in a bilayer as a function of helix sequence. There is a high degree of qualitative similarity (see Figure 3.32).

For a quantitative comparison it is difficult to compare ΔG values for similar systems, as

the only sequence which was investigated by both methods is M2-19; however, the ΔG_{app} and ΔG_{pmf} values for this sequence are very different owing to the bilayer distortion seen in the PMFs. This level of distortion is unlikely to occur in the experiments owing to flanking residues around the H-segment.

Despite this difficulty comparing identical quantities, the differences between ΔG_{app} and ΔG_{pmf} are large enough to be of significance. The ΔG_{app} values measured by experiment vary from $\sim$5 to $\sim$-7 kJ/mol. The value of ΔG_{pmf} for the M2-29 helix, by contrast, is -67 kJ/mol. Given the greater length of M2-29 compared with M2-19(-1), it might be expected that the ΔG_{pmf} for the former would be greater than the ΔG_{app} for the latter, but the observed difference of an order of magnitude is large enough to suggest that the methods are measuring different quantities.

The significant difference is not limited to comparisons between ΔG_{app} and ΔG_{pmf}; the additive model showed estimates of ΔG^{add} consistent with the PMF calculations for a range of hydrophobicity scales, with the exception of the estimate using the translocon-mediated scale. These estimates were of a much smaller magnitude. This effect has been observed and commented on [184], with a range of possible explanations put forward including lipid composition [186], protonation state [187], and density of protein in the membrane [188], with limited success.

The effect of bilayer width on ΔG_{pmf} suggests an alternative explanation for the 'compressed' nature of the ΔG_{app} values. Inserting a helix into a thinner hydrophobic region reduces the measured ΔG_{pmf}, showing that the free energy change when transferring a helix between different environments is heavily dependent on the hydrophobicity of these environments. This suggests that a close look at the individual steps involved in the *in vivo* insertion process may yield an explanation for the compression of ΔG_{app} values relative to other measures: that if translocon-to-bilayer partitioning is independent of water-to-bilayer partitioning, then ΔG_{app} would indeed be smaller than ΔG_{pmf} [184].

In the *in vivo* insertion process, a nascent chain is synthesised in the ribosome. This binds onto the Sec translocon complex with the assistance of the signal recognition particle (SRP). The nascent chain is then fed through the translocon pore by some mechanism which is not fully understood. Both power stroke [203] and Brownian ratchet [30] models have been proposed. Once the chain has entered the translocon pore, it folds [204] and then either partitions into the membrane, or translocates across the membrane. The mechanism for selecting which sections of the polypeptide chain may partition into the membrane is

not known. It may involve simple diffusion from the translocon pore into the membrane [205], or the hydrophobicity of the chain may control gating of the translocon [42]. We are therefore left with the following questions about the mechanism of the translocon:

1. How is the chain pushed out of the ribosome? How tightly is this bound to the chain entering the translocon?

2. How does the translocon gate open? Is it controlled by the peptide in the translocon pore or does it open transiently, perhaps as a result of thermal fluctuations?

3. Is partitioning from the translocon to the bilayer reversible?

4. What environment does the helix partition to?

The answers to these questions will affect our interpretation of ΔG_{app}. The mechanism by which the nascent chain is pushed into the translocon affects which part of the thermodynamic cycle ΔG_{app} measures. If the movement of the chain is only loosely coupled to the synthesis of the chain in the ribosome, then the chain would experience a water-filled environment before entering the translocon. If in addition insertion into the translocon occurs by an equilibrium or near-equilibrium process then ΔG_{app} may be equivalent to or dependent on ΔG_{wb} in Figure 3.33, with a further contribution from helix folding, presumably in the translocon. However, if the motion of the chain through the translocon is tightly coupled to protein synthesis in the ribosome, then insertion of the chain into the membrane will have effectively zero free-energy cost and ΔG_{app} would be equivalent to or dependent on ΔG_{trbi} in Figure 3.33.

Whether the helix partitioning from the translocon pore into the membrane is reversible also impacts on the interpretation of ΔG_{app}. Only if the partitioning is reversible will ΔG_{app} be related to ΔG_{trb}. If the translocon gate opening is substrate-controlled, as suggested by PMF calculations [42], one can imagine that the helix partitioning could be irreversible, governed by some conformational change in the translocon. If the gating is controlled by thermal fluctuations then the partitioning would presumably be by an equilibrium process. The rate of extension of nascent chains from the ribosome is of the order of milliseconds to seconds [38]. This seems sufficient time for a given helical segment to equilibrate between the translocon pore and the membrane; it is also consistent with the observation that the individual ΔG_{app} values derived from translocon-mediated experiments are well correlated with the Wimley-White hydrophobicity scale, and that this scale has a 95% accuracy for predicting membrane protein helices from sequence [206].

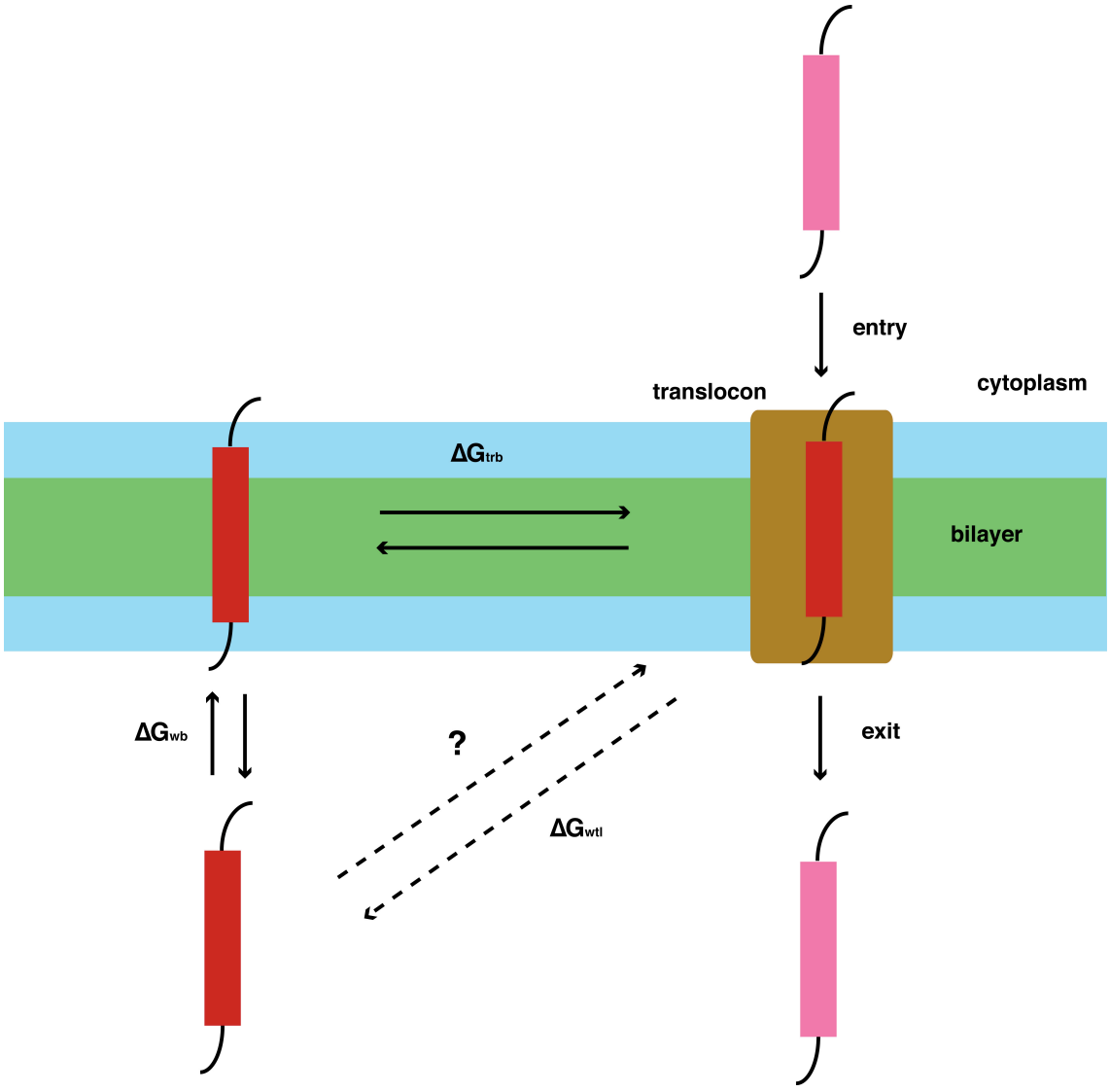


Figure 3.33: A schematic for helix insertion, outlining the possible relationships between translocon-mediated experiments and PMF calculations. The horizontal arrows (ΔG_{trb}) indicate a possible equilibrium between the peptide in the translocon and in the bilayer, which may be equivalent to the ΔG_{app} measured in translocon-mediated experiments. The vertical arrows (ΔG_{wb}) denote the free energy change between a folded helix in the bilayer and in water, the quantity measured by the CG PMF calculations. The diagonal arrows (ΔG_{wtr}) denote possible partitioning between a folded helix (red rectangle) in water and in the translocon. The entry and exit processes in the translocon may well be non-equilibrium, and involve peptides which have not fully folded (pink rectangles). Figure based on a similar schematic in [184].

The environment that the helix partitions to also affects the interpretation of ΔG_{app}; if the helix partitions directly into a lipid environment, then ΔG_{app} would be related to ΔG_{trb}, but if the helix is shielded from the lipids by, for example, helices of membrane proteins already present, then ΔG_{app} would be unrelated to ΔG_{trbi}. The suggestion that some helices partition into an environment other than lipids is supported by the observation that $\sim$25% of transmembrane helices in multispanning membrane proteins exhibit a predicted ΔG_{app} greater than 0 [36], though this may also be explained by some irreversible partitioning process. It is also supported by the observation that marginally hydrophobic helices can insert when surrounding hydrophobic regions are included [200].

So the ΔG_{app} values measured by experiment may represent a variety of quantities. If the chain moves into the translocon by a near-equilibrium process, and partitions into the membrane by a near-equilibrium process, ΔG_{app} will correspond to $\Delta G_{wtr} + \Delta G_{trb}$; if it moves into the translocon by a near-equilibrium process but partitions irreversibly then ΔG_{app} will correspond to ΔG_{wtr}. However, the observed discrepancy between ΔG_{app} values and a variety of other measurements of free energy insertion may be explained if we assume that the chain movement into the translocon is tightly coupled to protein synthesis in the ribosome; then it has effectively zero free energy cost. If helix partitioning into the membrane occurs by an equilibrium process then ΔG_{app} will correspond to ΔG_{trb}. ΔG_{app} will then depend on the relative hydrophobicities of the bilayer interior and the translocon interior; given the pore ring inside the translocon it seem reasonable to consider the translocon interior as an environment of intermediate hydrophobicity relative to the membrane and water. Measurements of ΔG_{pmf} on bilayers of varying width show that ΔG_{trbi} would be smaller than ΔG_{wb}, if the translocon interior was indeed of intermediate hydrophobicity. We saw that increasing the hydrophobic width by 1 nm increases ΔG_{pmf} by $\sim$37 kJ/mol for M2-39. Given the length dependence of ΔG_{pmf}, together with differences in lipid composition, protein concentration, protein folding states and the translocon pore environment it is perhaps better to regard this as a qualitative rather than quantitative estimate of the conversion factor for the translocon mediated experiments. However, it does show that the hydrophobic pore ring in the centre of the translocon is very likely to have a significant effect on measured ΔG_{app} values, and may help to explain some of the disparity between the Hessa hydrophobicity scale and a variety of others.

The effect of the translocon environment on free energies of insertion has been explicitly calculated using free energy perturbation on a polyleucine helix containing one arginine

residue [207]. This resulted in a ΔG_{ins} of 27 kJ/mol for arginine and -4.2 kJ/mol for leucine. These values show the same kind of compression as the Hessa scale, confirming that the transfer of residues from the translocon into the membrane produces ΔG_{ins} values with a smaller range than from water to a bilayer.

3.5 Conclusions

The 1-dimensional helix-bilayer PMFs shown in this chapter have shown that kinetic trapping may occur in some simulations. This may produce PMF profiles where certain degrees of freedom have not fully equilibrated, thus giving a PMF with higher energy than a fully equilibrated system would produce. Inspection of simulations and comparison with expected behaviour can correct this in some cases. Either PMFs restraining multiple degrees of freedom, or using multiple simulations with different starting positions, are required to fully describe the energy landscape. Restraining a particular degree of freedom, if there are concerns about it equilibrating properly, may solve the problem at the expense of reducing the amount of available information in a profile.

The significant differences in ΔG_{pmf} for M2-19, M2-29 and M2-39 show that helix length is significant for stability in a bilayer. Distortion caused by the short M2-19 helix meant that the IF state was favoured over the TM state, while the longer M2-39 helix was able to tilt and kink to maximise favourable interactions with lipid headgroups and tails. These results are consistent with studies of the effect of helix length using translocon-mediated insertion experiments [196].

The depth of the PMF profiles are consistent with simple estimates using a variety of hydrophobicity scales, including those derived from computational single-residue PMFs, continuum electrostatics calculations, water-cyclohexane partitioning and the prevalence of residues in the centre of proteins. However, the well depths and ΔG_{pmf} were generally much larger than ΔG_{app} values measured by translocon-mediated insertion experiments applied to similar sequences, a discrepancy that has been noticed using a variety of computational free energy measures. This is probably because ΔG_{app} values measure a quantity more complex than the free energy change of a helix from water into a bilayer. If the passive diffusion model of translocon function is correct, then ΔG_{app} may measure the free energy change on a helix moving from the translocon pore into the membrane. The dependence of ΔG_{pmf} values on the hydrophobic thickness of a bilayer suggests that the thickness of the pore ring in the translocon may contribute to explaining the disparity between computa-

tional hydrophobicities and the translocon-mediated insertion scale.

It has been demonstrated that a very similar PMF profile is obtained whether a 39 residue sequence is restrained as an ideal helix, or just the central 19 residues of the sequence. Attempting to calculate the PMF of a marginally hydrophobic helix produced a profile which seems unphysical.

Chapter 4

Multiscale Simulations of Eph Receptor Transmembrane Helices

4.1 Introduction

4.1.1 Receptor Tyrosine Kinases

Receptor tyrosine kinases (RTK) are high affinity cell surface receptors essential for the regulation of normal cell processes. Their structure includes an extracellular ligand binding domain, a single TM helix, and an intracellular tyrosine kinase domain [208]. They are classified according to their structural and ligand affinity properties into several families, including the epidermal growth factor (EGFR), fibroblast growth factor (FGFR), vascular endothelial growth factor (VEGFR), rearranged during transfection (RET) and erythropoietin-producing hepatocellular (Eph receptor) families [125]. RTKs conduct biochemical signals across the membrane via lateral dimerization in response to ligand binding [125]. Although the exact mechanism of this dimerization is not fully understood for all RTKs, the fact that the TM domains of all 58 human RTKs form stable dimers in a membrane suggests that these TM regions play an important role in RTK signalling [209].

4.1.2 Eph Receptors

Eph receptors are RTKs involved in a variety of development processes essential to the functioning of the cell, including guiding neuronal growth, cell morphogenesis, angiogenesis, tissue patterning and neural plasticity [210, 211, 212, 213]. They have also been implicated in a variety of cancers [214, 215, 216].

As with other RTKs, Eph receptors have an N-terminal extracellular region, contain-

ing the ligand binding domain; a single transmembrane helix; and an intracellular region containing the tyrosine kinase domain [217, 213]. Eph receptors are classified according to sequence familiarity and binding affinity as either A receptors or B receptors [211]. EphA receptors bind to ephrinA, which is bound to the membrane by a glycosylphosphatidylinositol (GPI) anchor. EphB receptors bind to ephrinB, which have a transmembrane domain followed by a short cytoplasmic region [218]. In vertebrates there are 10 EphA receptors (EphA1-EphA10) and 6 EphB receptors (EphB1-EphB6).

Eph receptors transduce a signal across the plasma membrane in response to binding of ephrin [219]. As both Eph receptors and ephrin are bound to the plasma membrane, most functionality stems from Eph receptors of one cell binding to ephrin on another cell (known as *in trans* interactions) [220]. There is some debate over the possible functionality of lateral (*in cis*) Eph-ephrin interactions [221]. Unusually, binding of Eph receptors to ephrin *in trans* triggers both 'forward' signals, into the Eph receptor-expressing cell, and 'reverse' signals, into the ephrin-expressing cell [222]. Both Eph receptors and ephrin are thought to be localised to lipid rafts [223, 224, 225], areas of the membrane enriched in cholesterol and glycosphingolipids.

Eph receptor binding to ephrin typically occurs with nanomolar affinity [218]. After binding, it is thought that the Eph-ephrin complexes form tetramers and possibly higher-order oligomers. Eph receptors not bound to ephrin bind laterally to the oligomers, presumably to extend the influence of the propagating signal [226, 227]. Interestingly, signals with different intensities can produce opposite effects [213].

4.1.3 EphA1 and EphA2

Structures of the dimeric versions of the transmembrane region of the EphA1 and EphA2 receptors have been obtained by solution NMR [140, 141]. The structures were obtained using 1:4 DMPC:DHPC bicelles, with subsequent refinement via MD simulations in a DMPC bilayer for EphA1 and a DPPC bilayer for EphA2.

The resulting EphA1 structure is a right-handed dimer (see Figure 4.1) with a crossing angle of $\sim -44°$, measured under acidic pH. The helices appear to interact via an N-terminal glycine zipper motif, $A^{550}X_3G^{554}X_3G^{558}$, with the separation between the helix axes measured at 0.62 nm. Under near-physiological pH conditions the Glu^{547} residue deprotonates, causing a rearrangement of the helix packing such that another conformation is observed with the dimer interacting via the C-terminal $A^{560}X_3G^{564}$ motif. This shall subsequently be

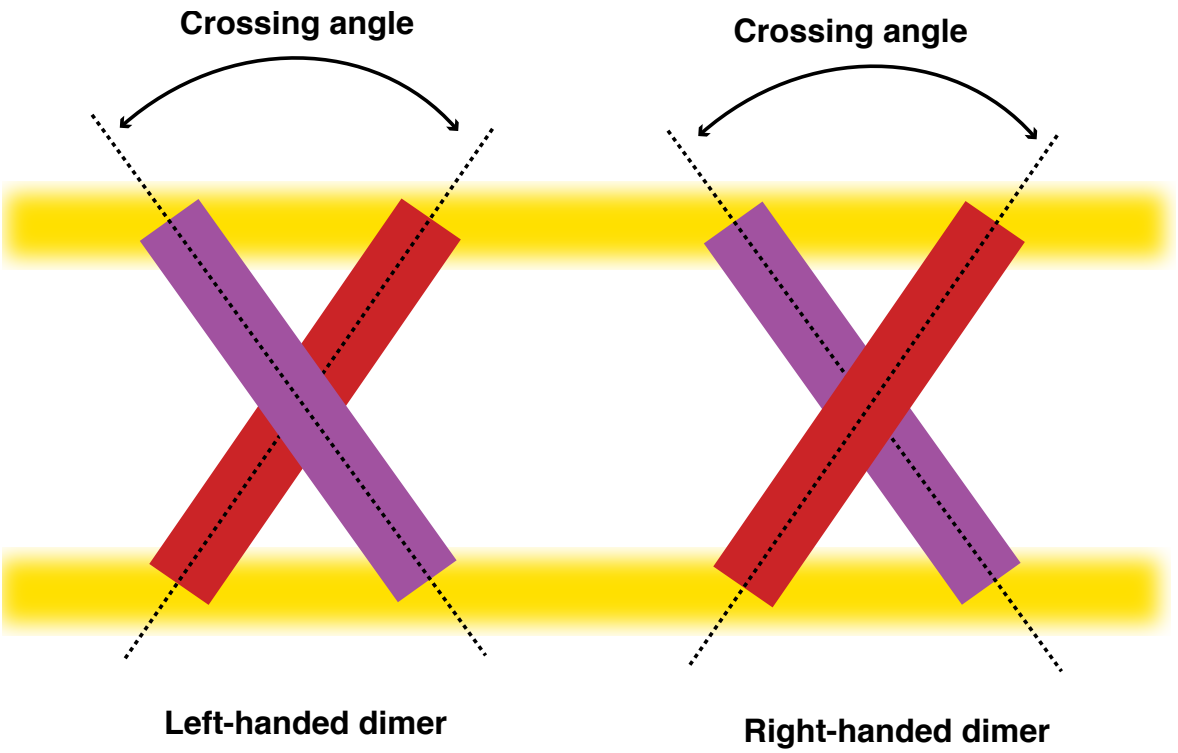


Figure 4.1: Schematic of left and right handed dimers. NMR structures show the TM region of EphA1 as a right-handed dimer, with the TM region of EphA2 as a left-handed dimer.

referred to in this thesis as the 'alanine' motif.

The EphA2 structure contains helices with a transmembrane region of similar length to the EphA1 helices, but adopts a left-handed dimer with crossing angle ∼15°. The resulting dimer is about 0.3 nm longer than the EphA1 dimer, so required a thicker bilayer for the MD refinement (DPPC vs DMPC). The dimer interacts via the heptad motif $L^{535}X_3G^{539}X_2A^{542}X_3V^{546}X_2L^{549}$ Interestingly, the EphA2 sequence contains an unused glycine zipper motif ($A^{536}X_3G^{540}X_3G^{544}$) similar to that observed in the EphA1 structure. A dimer formed from the interaction of the glycine zipper motif is expected to form a right-handed helix. Sample NMR structures of EphA2 and EphA1 are shown in Figure 4.2.

The dependence of the packing mode of EphA1 on pH has lead to the suggestion that its physiological role may be that of an extracellular pH sensor. It also provides evidence for EphA1 functioning by the rotation-coupled activator mechanism [228, 229], whereby switching between different dimeric packing states changes the receptor from an inactive to an active state. Similarly, the presence of both the glycine zipper and heptad motifs in the

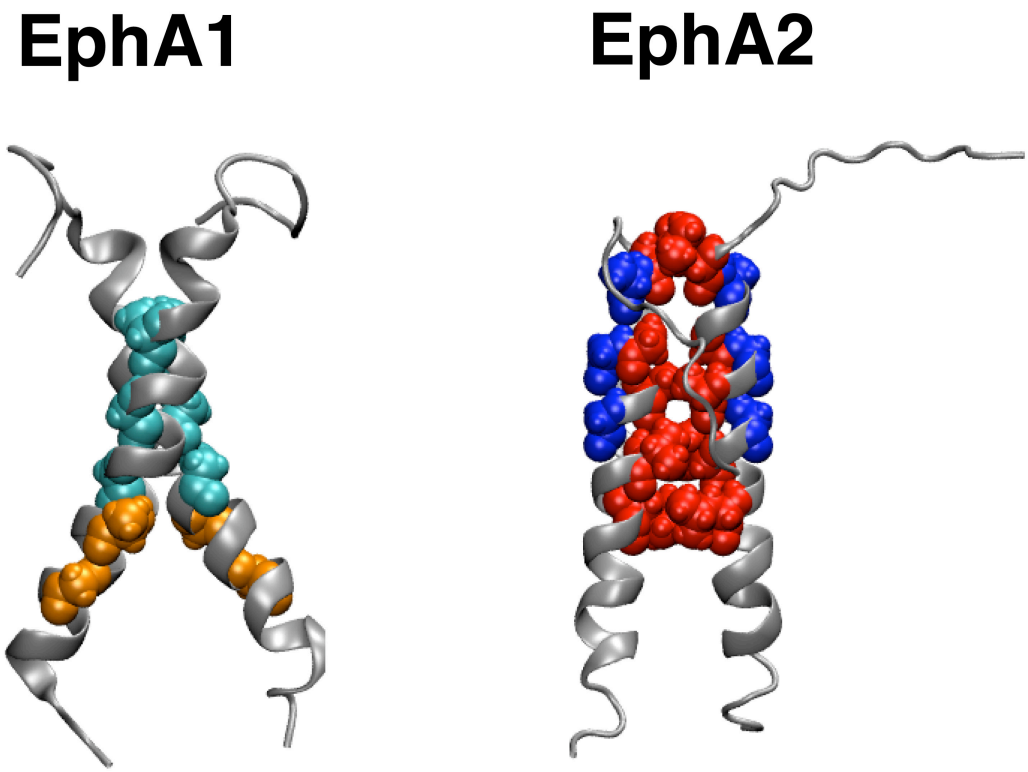


Figure 4.2: NMR structures of the TM regions of EphA1 and EphA2. The glycine zipper motif for EphA1 is shown in cyan, with the alanine motif in orange. The glycine zipper motif for EphA2 is blue, with the heptad motif in red. The biological unit contains an extracellular region (including the ephrin binding domain) and an intracellular region (including the juxtamembrane segment and kinase domain).

Table 4.1: Simulated EphA1 and EphA2 sequences

Helix	Sequence
EphA1	GLTG^{545}GEIVA^{550}VIFGL^{555}LLGAA^{560}LLLGI^{565}LVFRS^{570}RR
EphA2	SGNL^{535}AVIGG^{540}VAVGV^{545}VLLLV^{550}LAGVG^{555}FFIHR^{560}RR

EphA2 sequence raises the possibility of the rotation-coupled activation mechanism being employed by several Eph receptors. An alternative explanation for the presence of multiple possible interaction motifs in the TM regions of EphA1 and EphA2 is that they help in the formation of oligomers of Eph-ephrin complexes after cell-cell contact. This hypothesis would suggest that dimers of EphA2-ephrin may form via (say) the heptad motif, but higher-order oligomers may pack in the TM region via the glycine zipper motif. This assumes that oligomers form via TM interactions, which may not be valid due to steric interactions between the extra-cellular or intra-cellular domains of Eph receptors.

The two suggested packing modes for the TM regions of EphA1 and EphA2, together with the available NMR structures, make them an excellent sample system to investigate helix dimerization via CGMD simulations. The simulations aimed to answer the following questions:

1. Can simulations of the TM regions of EphA1 and EphA2 reproduce the structures seen from NMR spectroscopy?

2. Can simulations of the TM region of EphA2 show an interaction mode via the glycine zipper motif?

3. Can simulations of the TM region of EphA1 show an interaction mode via the alanine motif?

4.2 Methods

4.2.1 Sequences

The simulated helices contained 31 residues, including the suggested transmembrane regions of 21-23 residues together with 4-5 flanking residues at each end. The sequences are shown in Table 4.1.

4.2.2 Coarse-Grained Simulations

Five unrestrained CG simulations were run for the TM region of EphA1, and five for the TM region of EphA2. Atomistic helices were generated from the required sequence using the build function of the Pymol program [194]. This produces helices with ϕ and ψ angles of $\sim$-56° and $\sim$-47° respectively. Local code then generated a coarse-grained representation. The helices were aligned as required, then added to a simulation box with randomly placed lipid molecules. Water molecules were then added, along with sufficient counter-ions to ensure that the system was electrically neutral. The system was subjected to 1000 steps of l-bfgs energy minimisation. A 200 ns equilibration simulation was performed with the helices restrained, which allowed a bilayer to form around the helices. The production simulation was then performed. Repeat simulations were performed by changing the initial positions of the lipid molecules and the initial velocities of all particles.

All CG simulations in this chapter were performed using Gromacs version 4.5.3 [177]. The temperatures of the helix, lipids and solvent were each coupled separately using the Berendsen algorithm [153] at 323 K. The time step was 40 fs. The pressure was semi-isotropically coupled using the Berendsen algorithm at 1 bar with τ_P = 1 ps and a compressibility of 5×10^{-6} bar^{-1}.

4.2.3 Coarse Grained to Atomistic Conversion

The method of Stansfeld et al. [230] was used to convert CG structures to atomistic equivalents. The process replaces fragments of CG lipids with their closest match from a library of atomistic lipid fragments. The backbone of the protein is grown from CG Cα particles using the Pulchra algorithm [231], and side chains are added using Modeller [232] while preserving the original coordinates of the CG particles. After applying this process to the CG coordinates, the system was solvated and subjected to further energy minimisation. A 1 ns simulation with the helices restrained with a force constant of 1000 kJ/mol/nm^3 allowed the system to further equilibrate before a 100 ns simulation was subsequently performed.

The force field used for atomistic simulations was GROMOS96 43a2 [233]. These simulations used semi-isotropic pressure coupling with the Parrinello-Rahman barostat [157], while the Berendsen thermostat [153] was used to couple the temperature of the lipid, helices and solvent to an external bath at 323 K. The water model was SPC [234]. The LINCS algorithm [161] was used to constrain bond lengths, while the Particle Mesh Ewald (PME) method [158] was used to model long range (>1 nm) electrostatics interactions. A cutoff of

Table 4.2: Summary of 'ideal' EphA2 simulations.

Sim number	Dimerisation time (ns)	Box dimensions (nm)	Lipid numbers
1	170	9.5×9.5×18.2	282
2	820	9.5×9.5×18.0	286
3	1410	9.7×9.7×17.5	293
4	320	9.4×9.4×18.6	275
5	20	9.5×9.5×18.2	281

1 nm was used for van der Waals interactions.

4.2.4 Analysis

Clustering analysis of the simulations was performed on snapshots of the Cα particles of the helices taken every 8 ns of simulation time, using the NMRCLUST method of Kelley and Sutcliffe [235]; this clusters conformations into groups on the basis of the root-mean-square deviation (RMSD) of the Cα particles in the simulation. The representative frame is chosen as the one with the RMSD closest to the average of the cluster group with the largest number of frames. All other analysis was performed using Gromacs tools or locally written code on snapshots taken every 400 ps.

The helices were said to form a dimer when their centres of mass were closer than 1.2 nm. Once the helix centres of mass were closer than this they did not get further apart in any simulation.

4.3 Results

4.3.1 EphA2

Coarse-Grained Simulations

Five CG unrestrained simulations of the dimerisation of EphA2 were run, all for a duration of 2000 ns. The simulation details are included in Table 4.2. Box and bilayer sizes were similar for all simulations. The helices formed a dimer in all simulations, with the dimerisation time varying from 20 to 1410 ns. Snapshots of dimer formation from simulation 1 are shown in Figure 4.3.

The rotation angle between the helices is a useful description of the interaction behaviour of the helices during the simulations. For the EphA2 simulations this angle was defined as the angle between the vectors linking the glycine zipper and heptad motifs on each helix, in

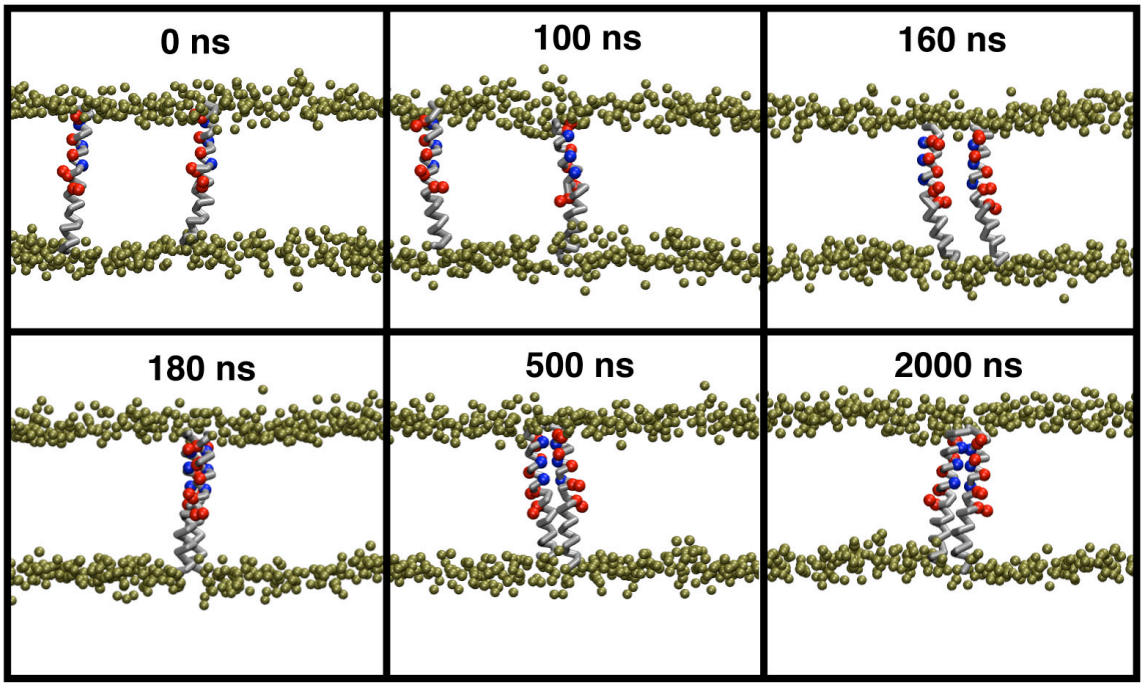
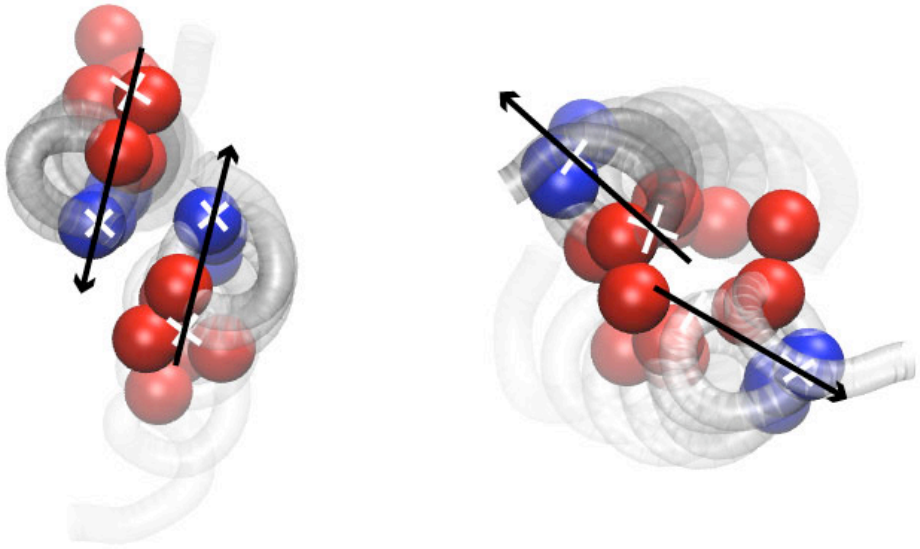


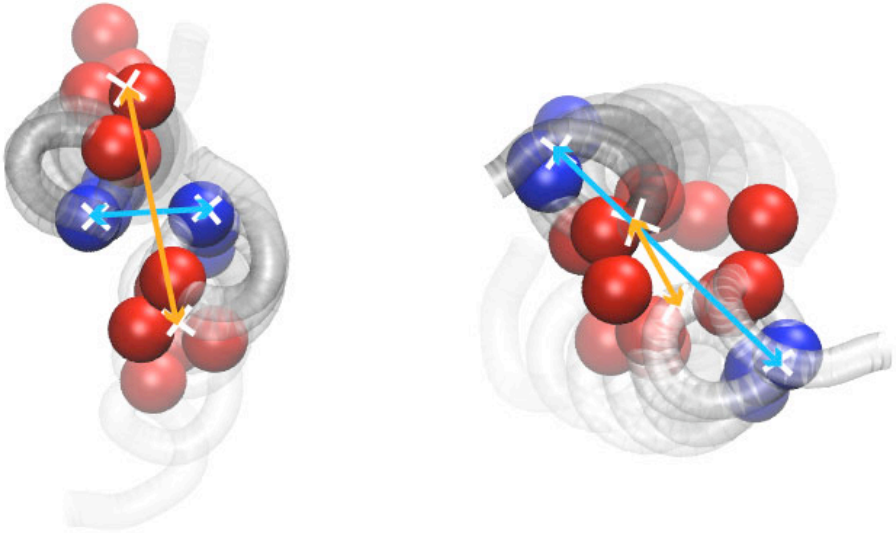
Figure 4.3: Snapshots from simulation 1 of EphA2. Helix backbones are silver, the glycine zipper motifs blue, heptad motifs red, and lipid phosphate and choline particles are tan. Lipid tails, waters and non-motif sidechain particles are omitted for clarity.

the plane of the bilayer, as shown in Figure 4.4 (a). The angle is calculated such that a rotation angle of -180° corresponds to the heptad motifs facing each other; +180° to the glycine zipper motifs facing each other; and 0° to the glycine zipper motif of one helix facing the heptad motif of the other helix, as shown in Figure 4.4 (b). Figure 4.5 shows the rotation angles as a function of time for each simulation. These rotation angles suggest that the helices adopt four distinct states in the simulations. The most common corresponds to a rotation angle of ∼150 to 180°, indicating interactions occurring via the glycine zipper motifs. This state will be referred to as state G_1, and occurs in simulations 1, 2, 3, and for much of simulation 5. A state corresponding to a rotation angle of ∼-150 to -180°, indicating interactions via the heptad motif, is seen in simulation 4. This will be termed state H_1. Two other states are seen in simulation 5, with rotation angles of ∼-100° and ∼80°. These states, termed H_2 and G_2 respectively, do not seem to be stable, and the helices switch between the two on a timescale of ∼10 ns for the first 600 ns. After ∼600 ns the system adopts state G_1, which remains stable for the rest of the simulation. Once in state G_1, the dimer remains stable in all simulations. One transition from state H_1 to H_2 was observed (in simulation 4), with the dimer switching back after ∼100 ns.

The rotation of the helices as a function of their separation is shown in Figure 4.6. The separation of the centres of mass of the helices ranges from 0.7-1 nm for state H_1, and from



(a) Vectors joining the centres of mass of the motifs, in the plane of the bilayer.



(b) The sign of the rotation angle depends on the separations of the centres of mass of the motifs.

Figure 4.4: Top view of the interaction of EphA2 helices. (a) Vectors are defined joining the centre of mass of the heptad and glycine zipper motifs (approximate centres of mass are shown as white crosses). The magnitude of the rotation angle is the angle between these vectors. (b) The sign of the rotation angle is defined by the relative distances between the motifs. If the distance between the glycine zipper motifs (light blue line) is less than the distance between the heptad motifs (orange line), the angle is positive. If the heptad motifs are closer, the angle is negative. Glycine zipper motifs are shown in blue, heptad motifs in red. Non-motif sidechain particles are not shown.

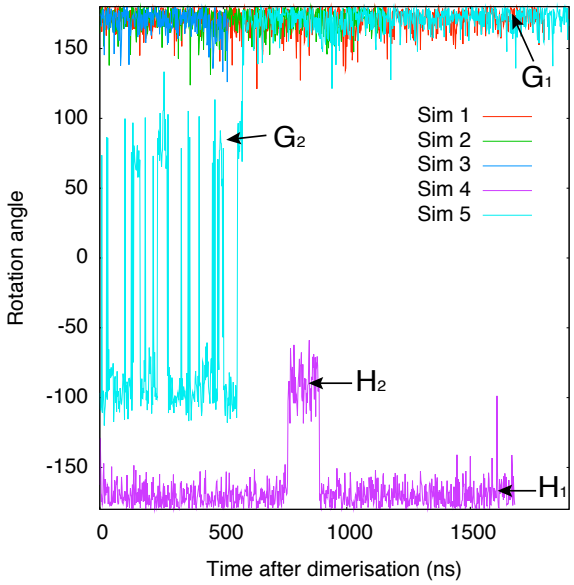


Figure 4.5: The rotation of the EphA2 'ideal' helices as a function of (dimerised) time for each simulation. Negative rotation angles correspond to the heptad motifs being closer; positive rotation angles correspond to the glycine zipper motifs being closer. A rotation angle of $\pm 180^\circ$ corresponds to the motifs directly facing each other.

0.8-1.1 nm for state G_1. The average centre of mass separation for the helices in simulations 1, 2 and 3 (state G_1) is 0.91 ± 0.09 nm (errors are standard deviations) but only 0.83 ± 0.08 nm for simulation 4 (state H_1). That the helices are closer in state H_1 might suggest that this is a more stable state than G_1. However, the fact that four simulations end in state G_1 and one ends in state H_1 suggests the opposite conclusion. The deviation from state H_1 to H_2 at around 800 ns in simulation 4 also suggests H_1 may not be as stable, though it is difficult to be certain of relative stability of different states from small numbers of unrestrained simulations.

Cluster analysis was used to produce the most representative structure from each simulation. The method chooses the RMSD of the Cα particles which is closest to the average in the simulation [235]. A clustered snapshot from each simulation is shown in Figure 4.7. This confirms that the most representative state for simulations 1, 2, 3 and 5 is interaction via the glycine zipper motif (G_1), while for simulation 4 it is interacting via the heptad motifs (H_1). However, both state G_1 and H_1 seem to produce right-handed dimers, although the NMR structure of EphA2 shows the helices in a left-handed dimer.

The average distance between each particle in the dimer part of simulations 1 and 4, corresponding to states G_1 and H_1 respectively, is shown in Figure 4.8. The symmetry of these matrices suggests that the helices are not displaced vertically with respect to one another in either state, confirming that the two motifs interact directly with each other as well as facing

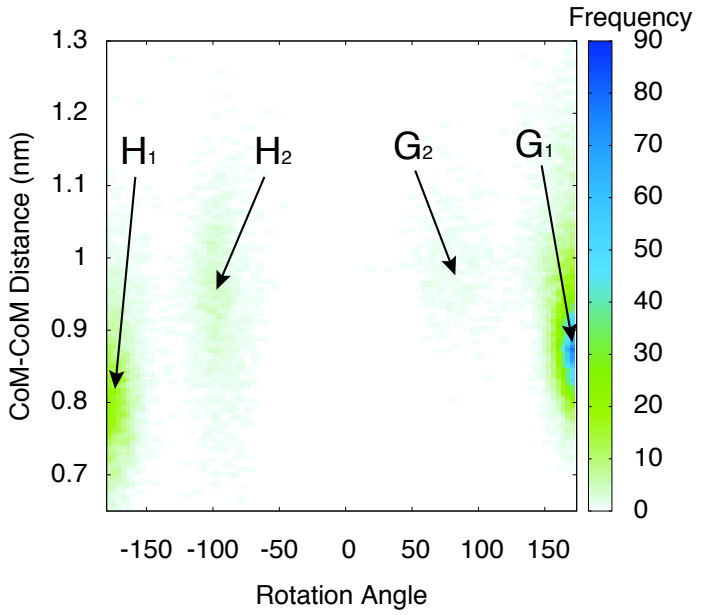


Figure 4.6: The rotation of the EphA2 'ideal' helices as a function of the separation of their centres of mass. Negative rotation angles correspond to the heptad motifs being closer; positive rotation angles correspond to the glycine zipper motifs being closer. A rotation angle of $\pm 180^\circ$ corresponds to the motifs directly facing each other.

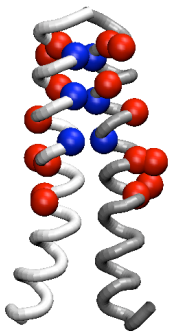
each other.

Atomistic Simulations

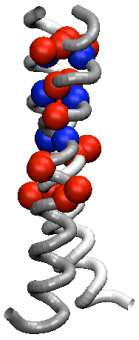
To check the relative stability of the glycine zipper-facing and heptad-facing structures, two sample structures were converted to an atomistic representation and a 100 ns simulation was performed. The structures chosen were the last frames of simulations 2 and 4, representing states G_1 and H_1 respectively.

The helix-helix centre of mass separation, the root mean-square deviation (RMSD) of the Cα particles and the solvent-accessible surface (SAS) of the two atomistic simulations are shown in Figure 4.9. The RMSD is a measure of how far the Cα particles deviate from their initial positions. SAS is a measure of how much of the helices is exposed to the lipid bilayer.

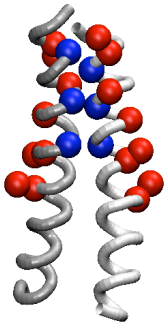
In both simulations the initial dimer structure appears unstable. For the glycine zipper-facing simulation, the helix separation rapidly increases from $\sim$1 nm up to $\sim$1.2 nm. This separation remains constant for about 60 ns, before increasing to $\sim$1.4 nm at the end of the simulation. This suggests that the helices are dissociating; a conclusion confirmed by the fact that the RMSD and the SAS steadily increase during the simulation. For the heptad-facing



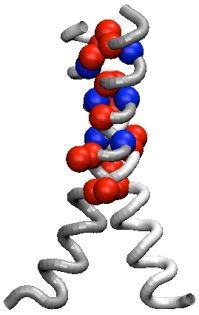
(a) Simulation 1



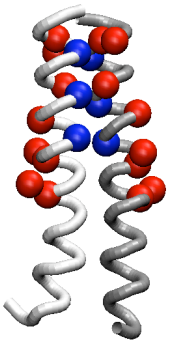
(b) Simulation 1 (side view)



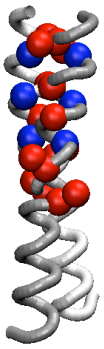
(c) Simulation 2



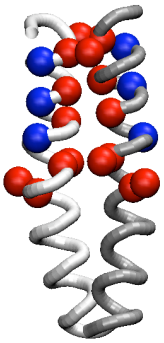
(d) Simulation 2 (side view)



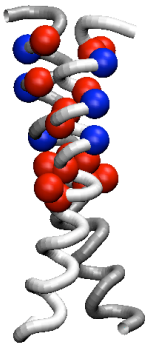
(e) Simulation 3



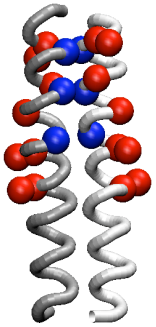
(f) Simulation 3 (side view)



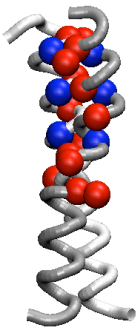
(g) Simulation 4



(h) Simulation 4 (side view)

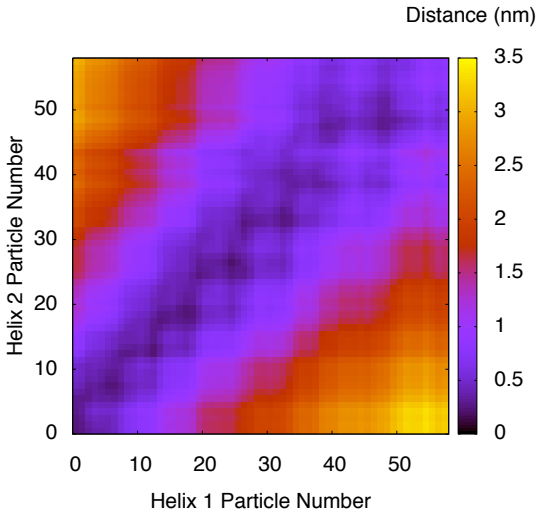


(i) Simulation 5

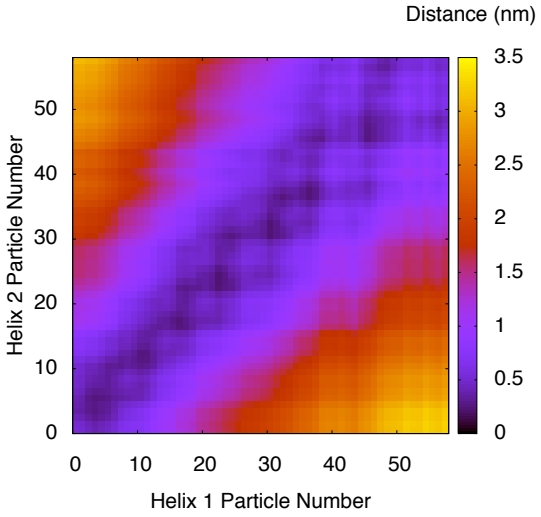


(j) Simulation 5 (side view)

Figure 4.7: Snapshots from the independent unrestrained simulations of 'ideal' EphA2 helices. These are the most representative of the dimer structure according to cluster analysis. The backbone particles are shown in white/grey. The heptad motifs are coloured red, the glycine zipper motifs blue. Waters, lipids and non-motif sidechain particules are omitted for clarity.



(a) Simulation 1

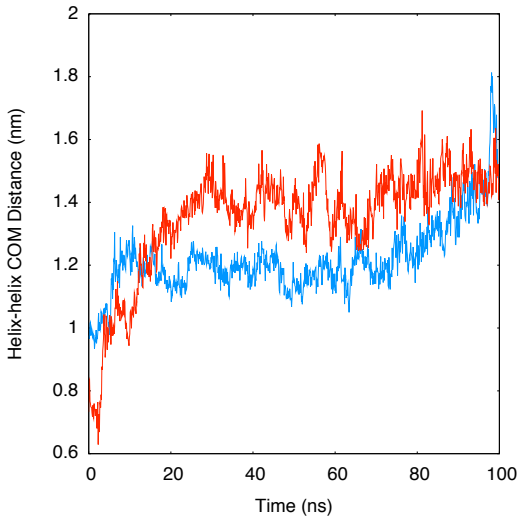


(b) Simulation 4

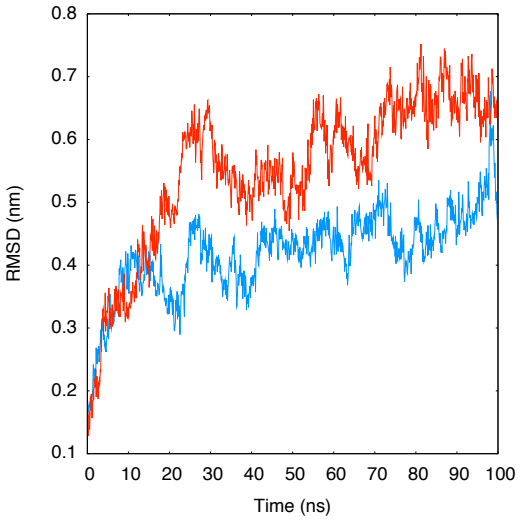
Figure 4.8: Average distance between each particle for the dimer part of simulations 1 and 4 for the EphA2 'ideal' helices, corresponding to states G_1 and H_1 respectively. The matrices are nearly symmetric, suggesting that the helices are not shifted vertically with respect to each other.

simulation, the helix-helix separation increases even more rapidly, reaching a seemingly steady value of ∼1.4 nm. The RMSD for the heptad-facing simulation is higher than for the glycine zipper simulation, although the SAS seems fairly similar for the two simulations.

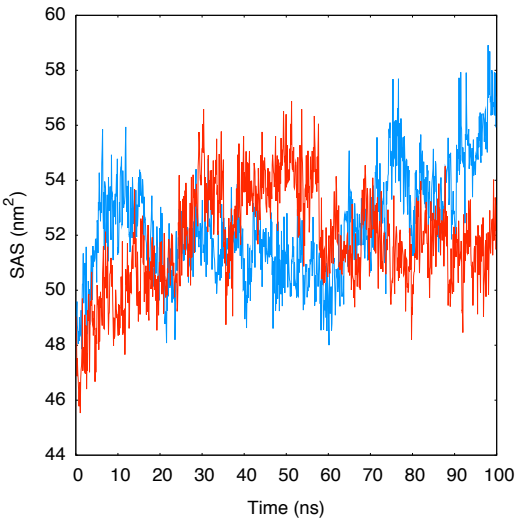
The snapshots from the glycine zipper-facing simulation are shown in Figure 4.10. These show that by 10 ns the helices are no longer interacting via the glycine zipper motif, and by 100 ns have separated to a significant distance. Similarly snapshots from the heptad-facing simulations (Figure 4.11) show that by 25 ns the helices are no longer interacting via the heptad motifs, and by 100 ns the helices have both tilted and separated significantly.



(a) Helix separation



(b) RMSD



(c) SAS

Figure 4.9: Analysis of EphA2 atomistic simulations. (a) Centre of mass separation of the helices. (b) Root mean squared deviation, a measure of how far the Cα particles move from their positions in the starting structure. (c) Solvent accessible surface, a measure of how exposed the helices are to their environment. The glycine zipper facing dimer is in blue, the heptad facing dimer is red.

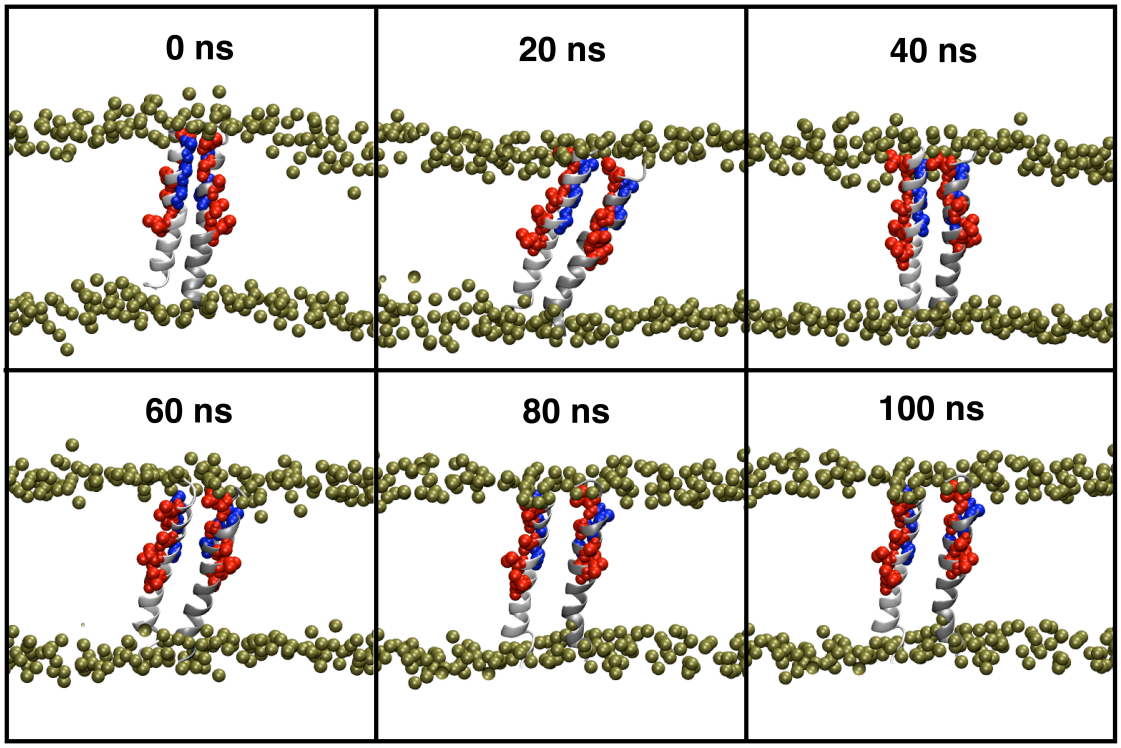


Figure 4.10: Snapshots from the EphA2 atomistic simulation starting with the glycine zipper motif facing. Waters, lipids (except phosphate group) and non-motif sidechains are omitted for clarity. Heptad motif is shown in red, glycine zipper motif in blue.

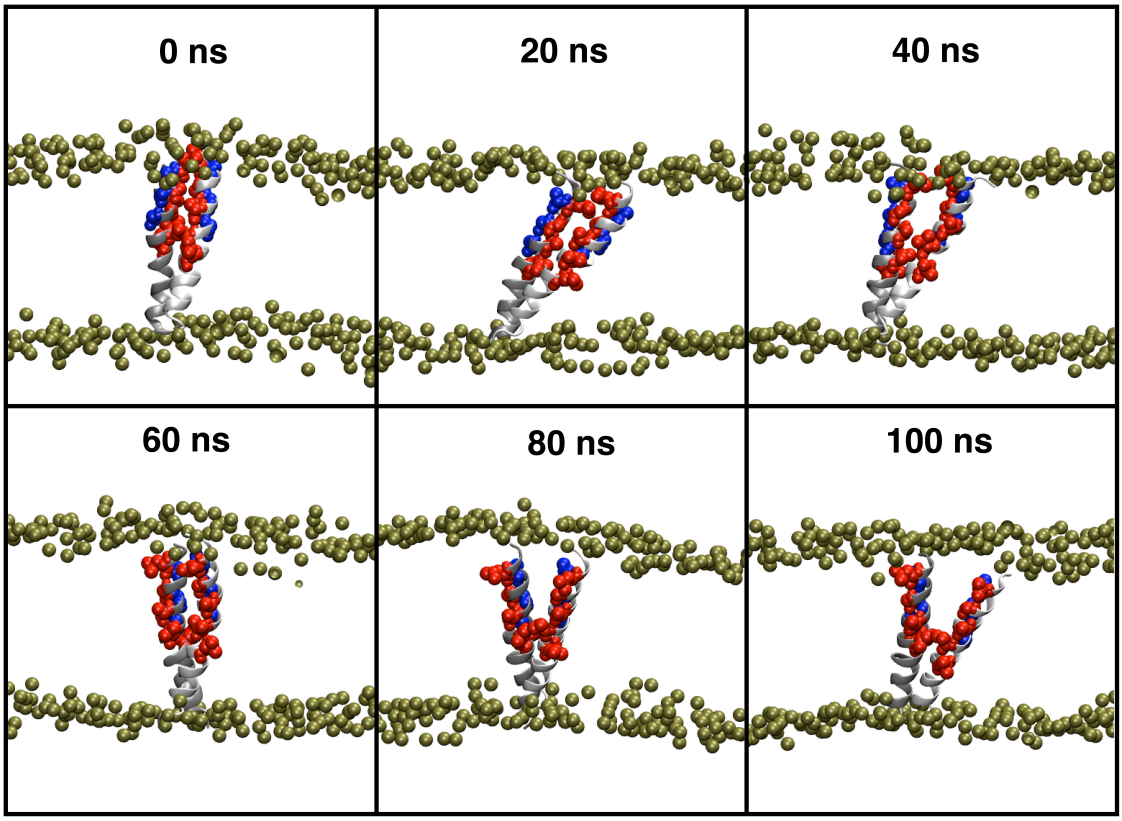
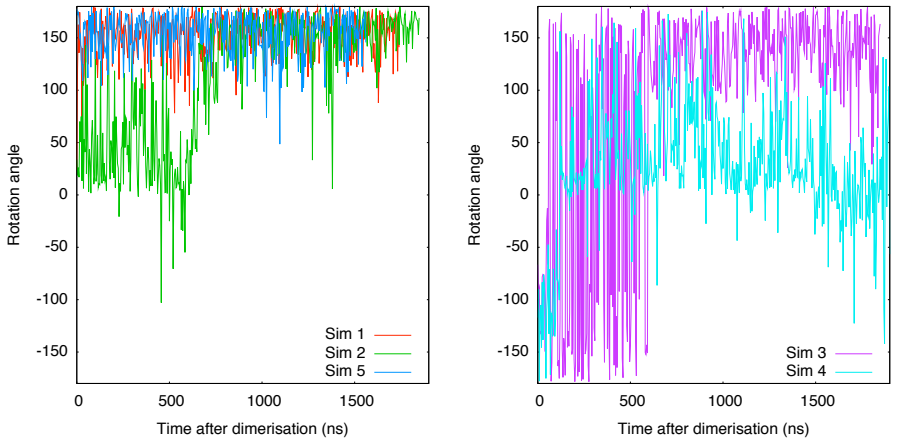


Figure 4.11: Snapshots from the EphA2 atomistic simulation starting with the heptad motif facing. Waters, lipids (except phosphate group) and non-motif sidechains are omitted for clarity. Heptad motif is shown in red, glycine zipper motif in blue.

Table 4.3: Summary of 'ideal' EphA1 simulations.

Sim number	Dimerisation time (ns)	Box dimensions (nm)	Lipid numbers
1	260	9.6×9.6×17.7	285
2	150	9.3×9.3×18.7	271
3	150	9.4×9.4×18.3	277
4	70	9.5×9.5×18.1	282
5	440	9.3×9.3×18.9	268



(a) Simulations 1, 2 and 5 (b) Simulations 3 and 4

Figure 4.12: The rotation of the EphA1 'ideal' helices as a function of (dimerised) time for each simulation. Negative rotation angles correspond to the alanine motifs being closer; positive rotation angles correspond to the glycine zipper motifs being closer. A rotation angle of $+180°$ corresponds to the glycine zipper motifs directly facing each other.

4.3.2 EphA1

Table 4.3 summarises the details of five unrestrained CG simulations of the EphA1 TM helices . The duration of each simulation was 2000 ns. A dimer was formed in each simulation, with the dimerisation time varying from 70-440 ns. The average dimerisation time was 214 ns, which compares with the average from EphA2 of 548 ns.

Figure 4.12 shows the rotation angles of the helices as a function of time. Similarly to EphA2, a rotation angle of $+180°$ corresponds to the glycine zipper motifs facing each other; $-180°$ to the alanine motifs facing each other. In simulations 1 and 5 the helices interact stably via the glycine zipper motifs. In simulations 2 and 3 the helices also adopt a stable interaction via the glycine zipper motifs, after sampling a variety of unstable states for ∼700-800 ns. In simulation 4 the helices seem to rotate relative to each other for the duration of the simulation, without adopting a stable position at any stage.

Figure 4.13 shows the rotation of the helices as a function of their separation. This also

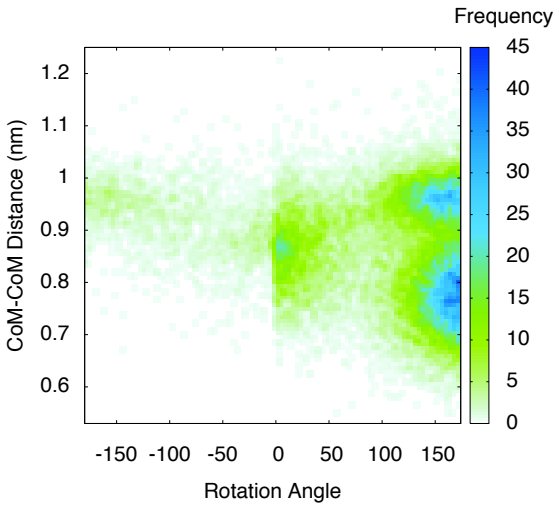
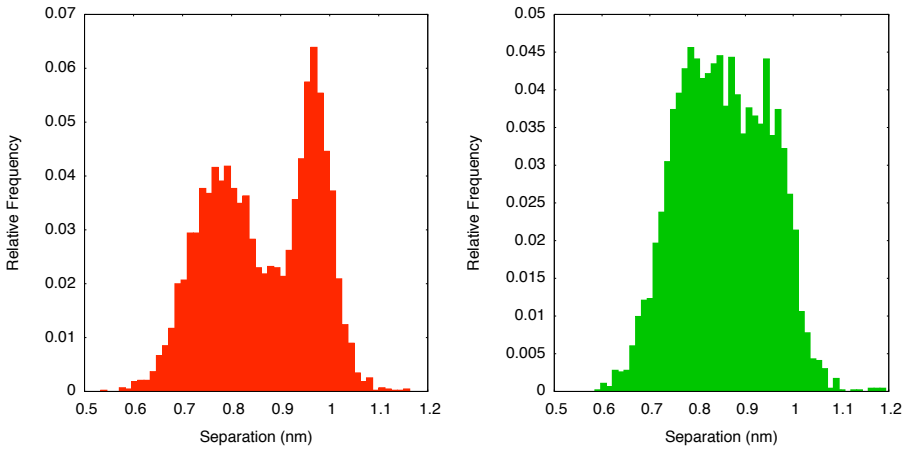


Figure 4.13: The rotation of the EphA1 'ideal' helices as a function of their separation. Negative rotation angles correspond to the alanine motifs being closer; positive rotation angles correspond to the glycine zipper motifs being closer. A rotation angle of $\pm 180^{\circ}$ corresponds to the motifs directly facing each other.

suggests that the interaction via the glycine zipper motifs, corresponding to a rotation angle of $+180^{\circ}$, is preferentially adopted, but shows this interaction taking place at two separate helix separations - one centered around a separation of 0.9-1 nm, the other around a separation of 0.7-0.8 nm. Histograms of the helix separation confirm that in simulations 1, 3 and 5 the helices seem to adopt two different states characterised by separations of $\sim$0.8 nm and $\sim$1 nm (Figure 4.14). This suggests that the helices may switch within these simulations between a 'close' and a 'distant' state. However, Figure 4.15 shows the helix centre of mass separation as a function of time for simulations 1 and 5, which show the helices switching rapidly between separations of $\sim$0.8 and $\sim$1 nm. This suggests that the varied separations may represent natural fluctuations together with some kind of steric barrier preventing much sampling at a separation of $\sim$0.9 nm.

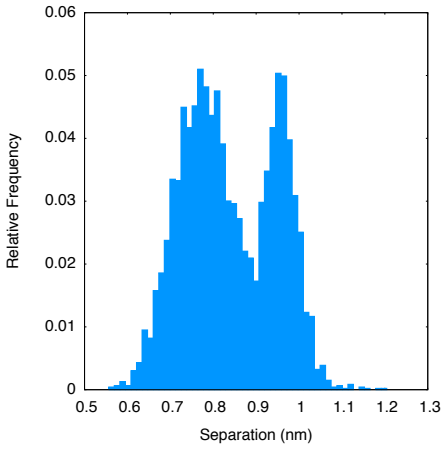
Cluster analysis was used to produce representative snapshots from each simulations, showed in Figure 4.16. These show the helices interacting via the glycine zipper motif for simulations 1, 2, 3 and 5. In simulation 4 the glycine zipper motifs on each helix are pointing in the same direction, rather than at each other, so different faces of each helix are interacting. The snapshots for simulations 1, 3 and 5 all show the helices in the 'near' state.

The average distance between each particle in the dimer part of simulations 4 and 5 is shown in Figure 4.17. The near-symmetry of the matrices shows that the helices are not vertically displaced with respect to one another.

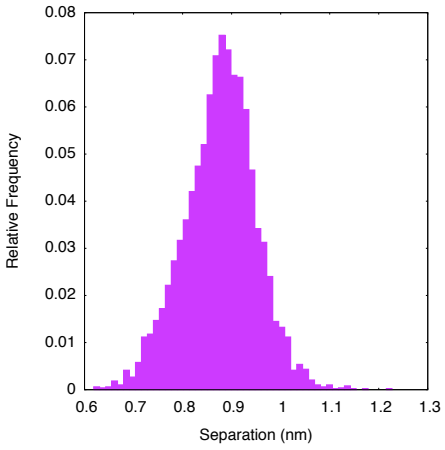


(a) Simulation 1

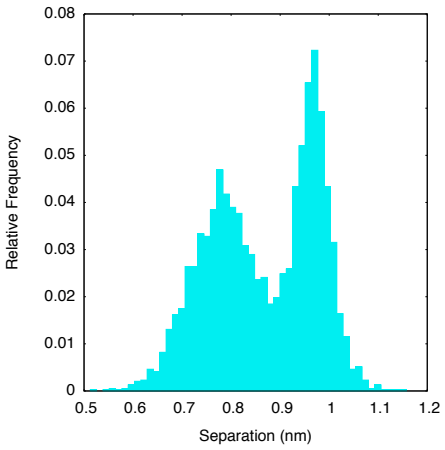
(b) Simulation 2



(c) Simulation 3

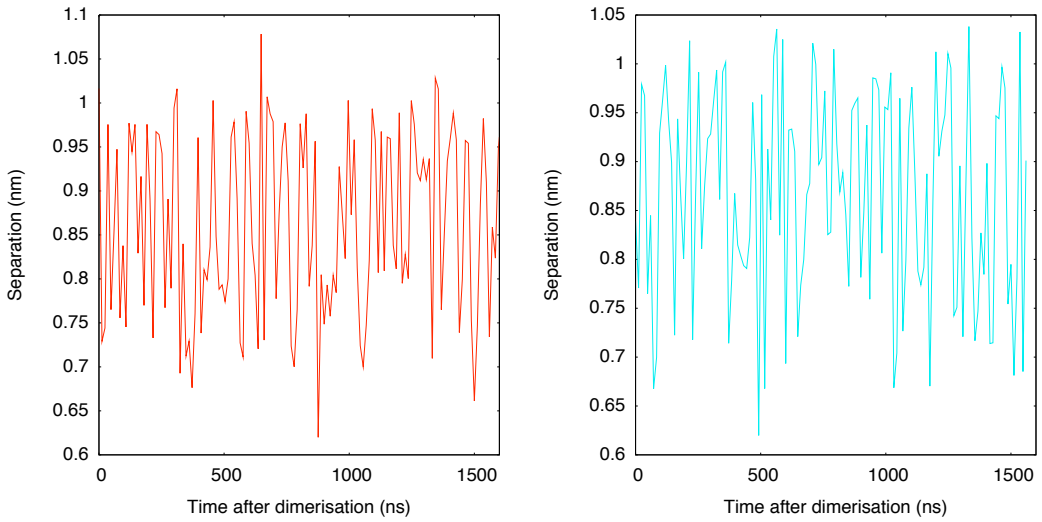


(d) Simulation 4



(e) Simulation 5

Figure 4.14: Histograms of the helix centre of mass separation in the five unrestrained 'ideal' EphA1 simulations.



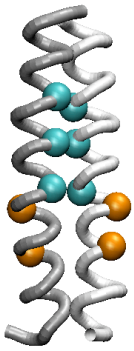
(a) Simulation 1 (b) Simulation 5

Figure 4.15: Helix centre of mass separation for simulations 1 and 5 of EphA1. These show the helices moving rapidly between separations of 0.7 and 1 nm.

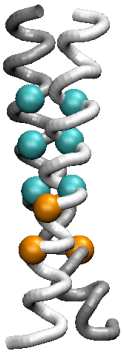
Atomistic Simulations

Similarly with the EphA2 simulations, the stability of the CG structure was tested by converting it to an atomistic equivalent. The final snapshot from simulation 1 was converted to an atomistic representation and a 100 ns simulation was performed. The helix-helix centre of mass separation, the root mean-square deviation (RMSD) of the Cα particles and the solvent-accessible surface (SAS) of the two atomistic simulations are shown in Figure 4.18.

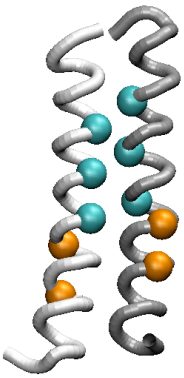
In contrast to the EphA2 atomistic simulations, the EphA1 helices maintain a dimer structure throughout the atomistic simulation. The average distance between the centres of mass of the helices varies between $\sim$0.7 - 1 nm. The distribution of these separations was clustered around a single value ($\sim$0.85 nm), rather than the two 'near' and 'distant' separations observed in the CG simulations (data not shown). Both the RMSD and SAS of the dimers increased during the simulation, suggesting that they were adjusting through the simulation, although snapshots from the simulation (Figure 4.19) do not show an obvious change in structure. It is possible that longer simulations, or repeats with slightly different initial structures or velocities, could show the helices dissociating. Significantly longer simulation times and/or large numbers of repeats would be needed to confidently assert that the atomistic structure is indeed stable.



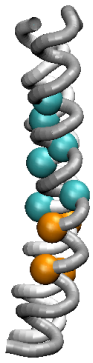
(a) Simulation 1



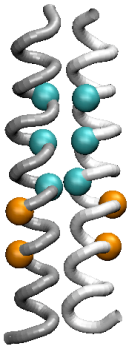
(b) Simulation 1 (side view)



(c) Simulation 2



(d) Simulation 2 (side view)



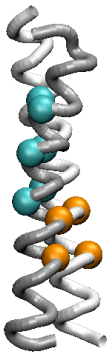
(e) Simulation 3



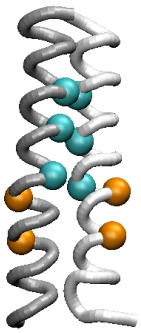
(f) Simulation 3 (side view)



(g) Simulation 4



(h) Simulation 4 (side view)

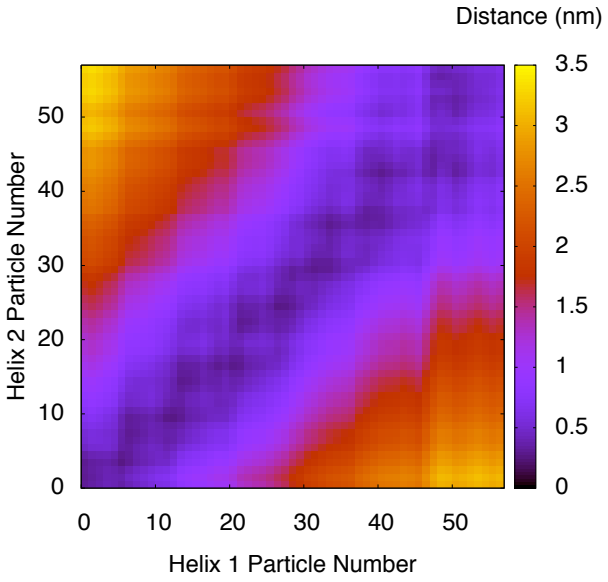


(i) Simulation 5

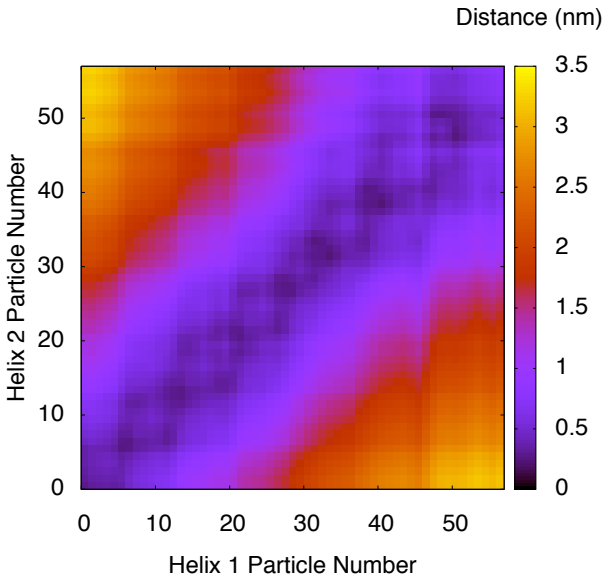


(j) Simulation 5 (side view)

Figure 4.16: Most representative structures from the independent unrestrained simulations of 'ideal' EphA1 helices. The backbone particles are shown in white/grey. The glycine zipper motifs are coloured cyan, the alanine motifs yellow. Waters, lipids and non-motif side chains are omitted for clarity. The snapshots for simulations 1, 3 and 5 show the dimers in the 'near' state.

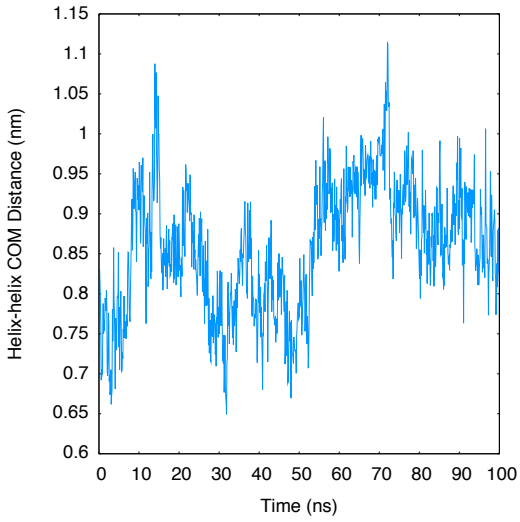


(a) Simulation 4

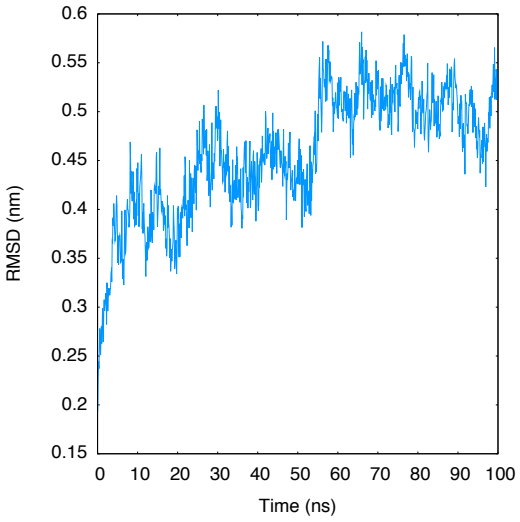


(b) Simulation 5

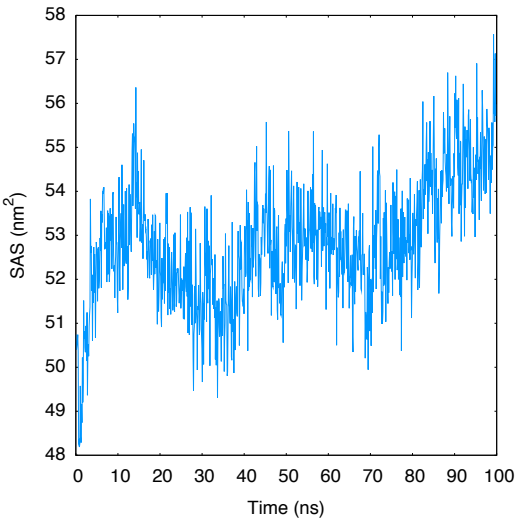
Figure 4.17: Average distance between each particle for the dimer part of simulations 4 and 5 for the EphA1 'ideal' helices. The distances suggest that the helices are not shifted vertically with respect to each other. They also suggest that the closer interaction is towards the N-terminal (which contains the heptad and glycine zipper motifs) rather than C-terminal end of the helices.



(a) Helix separation



(b) RMSD



(c) SAS

Figure 4.18: Analysis of atomistic EphA1 simulations. (a) Centre of mass separation of the helices. (b) Root mean squared deviation, a measure of how far the Cα particles move from their positions in the starting structure. (c) Solvent accessible surface, a measure of how exposed the helices are to their environment.

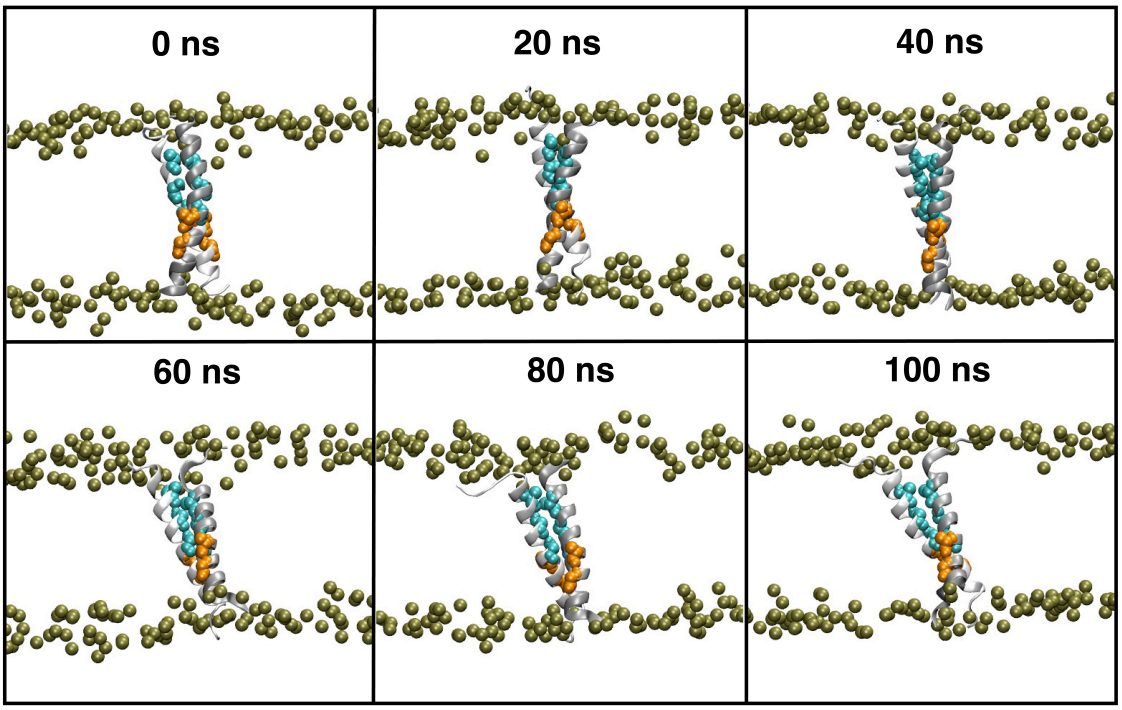


Figure 4.19: Snapshots from the atomistic simulation of EphA1. Waters, lipids (except phosphate group) and non-motif sidechains are omitted for clarity. Glycine zipper motif is shown in cyan, alanine motif in orange.

4.4 Discussion

4.4.1 Dimerising Time

The average dimerisation time for EphA1 was 214±57 ns, compared with 548±227 ns for EphA2 (errors are standard errors). We might expect dimerisation to be random, caused by stochastic collisions between helices driven by simple diffusion; alternatively, long range effects such as hydrophobic mismatch, or correlated lipid motions, could help or hinder the dimerisation process. To investigate the significance of the dimerisation times for the Eph receptor helices, a simple model of 2-dimensional diffusion was constructed to estimate how quickly two helices would come together if they moved randomly. In the model two points move a distance of 0.01 nm every timestep across a square of 9.5×9.5 nm^2, subject to periodic boundary conditions. The timestep was parameterised by comparing the mean squared distance (MSD) travelled in the model with the MSD of two separate helices in a CGMD simulation. The time at which the points came within 1.5 nm of each other - corresponding to helices colliding - was recorded. This was repeated 1000 times. The average time of collision was 66 ns. This is significantly less than the collision time observed in the simulations, which might suggest that the helices do not associate every time they collide. However, helix-helix

distances suggest that both the EphA1 and EphA2 helices associate every time they reach a separation of ∼1.5 nm. This suggests that some kind of correlated lipid motion, or annular lipids around the helices, may prevent the helices from colliding as frequently as if they were moving randomly.

4.4.2 EphA2 Interaction Modes

CG simulations of the EphA2 TM helices suggest that they may interact with both glycine zipper and heptad motifs, although both seem to be right-handed structures, in contradiction to the NMR structure [141]. The fact that four simulations end in state G_1, and only one in state H_1, suggests that the former state is more energetically favourable. This is supported by the fact that once in state G_1 the dimer stays in state G_1 in all simulations, but when in state H_1 a temporary transition to H_2 was observed. That G_1 is more stable than H_1 is somewhat surprising as the NMR structure shows the helices interacting via the heptad motif.

The structure of the helices interacting via the heptad motifs obtained by CG simulation can be compared directly with the NMR structure [141] (Figure 4.20). Both structures show the helices interacting via the heptad motifs; however, the NMR structure is a left-handed dimer, while the CG structure is a right-handed dimer. This therefore suggests that the favourable interactions which form the left-handed dimer *in vivo* can be reproduced in the CG simulations, but the detailed packing is not reproduced. The different handedness of the dimers cannot be easily reconciled by simply by considering the different hydrophobic thickness of the NMR and CG simulation environments (DMPC/DHPC bicelle and DPPC bilayer respectively).

Conversion of frames representing both G_1 and H_1 to an atomistic representation produced simulations where the dimer structures were not stable; in fact, the helices appear to be dissociating during the 100 ns of the simulation. This observation needs to be treated with caution, as this represents the results of just two simulations using a single force-field and a single lipid type. However, assuming that the observation is repeatable, there are two apparent explanations. The first is that the CG model may produce dimer packing which is not representative of the *in vivo* packing of EphA2. This is supported by the fact that the CG model was unable to reproduce a left-handed dimer similar to the NMR structure. However, the CG model used has been shown to reproduce the behaviour and packing of dimers such as GpA and syndecan [135, 136]. The second explanation for the instability of the EphA2

atomistic dimer structures lies in the environment of the dimer system. Characteristics such as lipid type and charge, bilayer thickness and protein concentration can have a significant effect on both the function of membrane proteins [236] and the stability of TM helix dimers [237]. The fact that Eph receptors are thought to be found in lipid rafts is another indication that the membrane environment may play an active role in Eph receptor dimerisation and stability [213]. Given that the NMR structure of the TM region of EphA2 was obtained in a different environment from that in the CG and atomistic simulations, it is therefore perhaps not surprising that differences in dimer packing were observed. These results underline the necessity of using a range of simulation techniques to make a confident prediction about the behaviour of a system. Validating simulation results with a variety of other measurements, ideally from experiment, is also crucial to maintain confidence in a simulation model. In the case of CG simulations of the TM Eph receptors, a range of simulations with a variety of different lipid types, at both CG and atomistic level, could provide a more thorough understanding of the effect the membrane environment plays on the TM helix packing.

4.4.3 EphA1 Interaction Modes

Unrestrained CG simulations of the EphA1 TM helices show the helices primarily interacting via the glycine zipper motif, with no evidence of the helices interacting via the alanine motif. There seem to be two states sampled within the simulations, a 'near' state with a helix centre of mass separation of $\sim$0.7 to 0.8 nm, and a 'distant' state with a helix separation of $\sim$0.9 to 1 nm. A subsequent atomistic simulation of the EphA1 helices suggests that the dimer remains stable, in contrast to the EphA2 helices, although the atomistic simulations provide no evidence of a distinction between the 'near' and 'far' states.

The structure of the helices interacting via the glycine zipper motif can be compared directly with the NMR structure [140] (Figure 4.21). Both structures are right-handed helices interacting via similar residues; however, the NMR structure shows a much larger crossing angle than the dimer derived from CG simulations. The crossing angle is heavily influenced by the hydrophobic width of the environment; the fact that the CG simulations were performed in a DPPC bilayer and the NMR structure came from a DMPC/DHPC bicelle, with energy relaxation by MD in a DMPC bilayer, may provide an explanation for the observed difference in crossing angles.

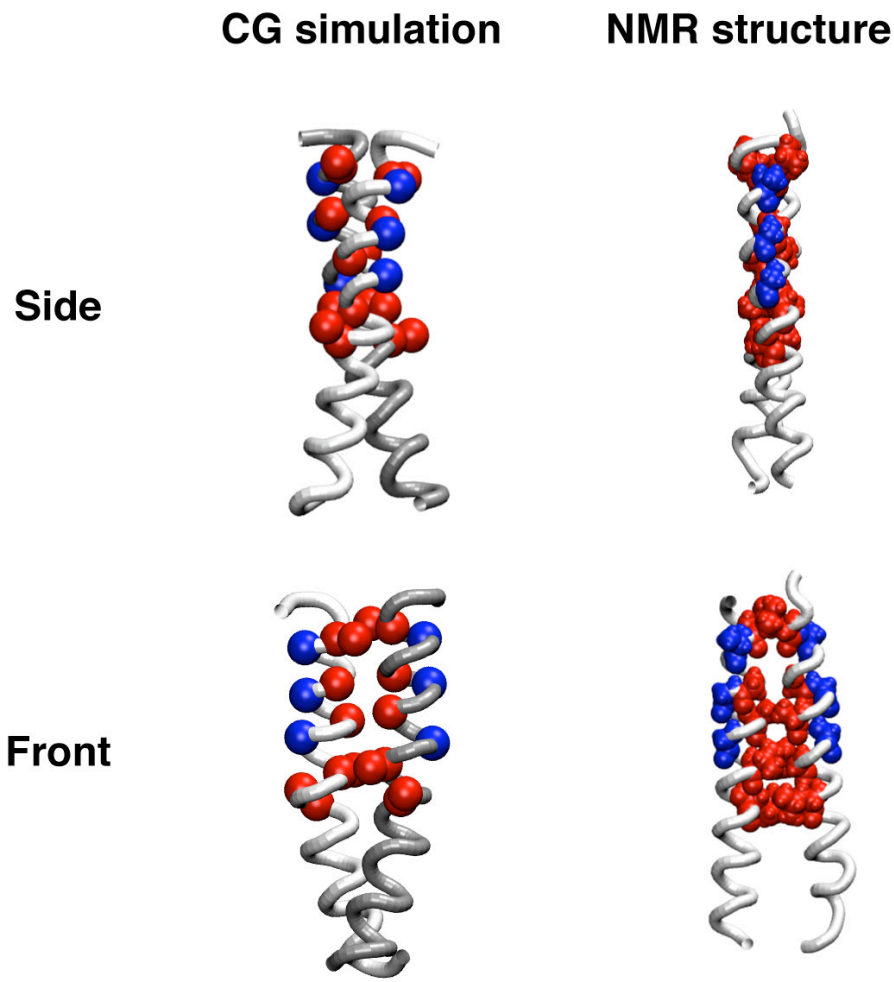


Figure 4.20: A comparison of the NMRCLUST snapshot from simulation 1 of the EphA2 simulations with the NMR structure from Bocharov et al. [141] (pdb code 2K9Y). Heptad motifs are shown in red, glycine zipper motifs in blue.

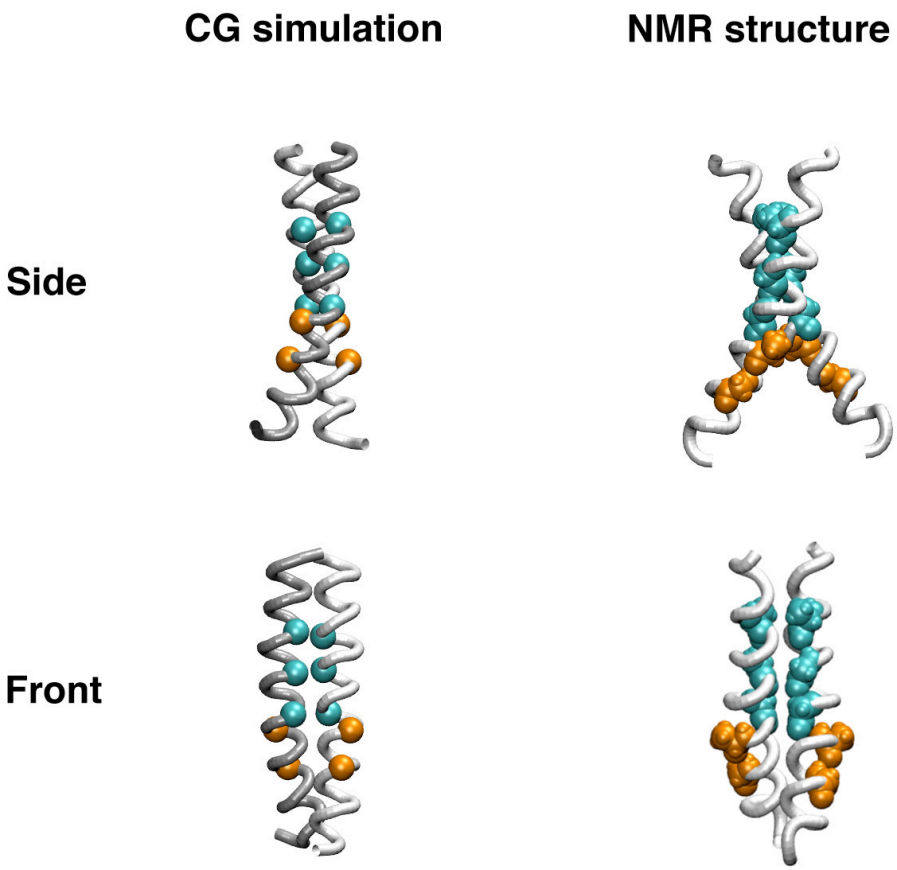


Figure 4.21: A comparison of the NMRCLUST snapshot from simulation 1 of the EphA1 simulations with the NMR structure from Bocharov et al. [140] (pdb code 2K1L). Glycine zipper motifs are shown in cyan, alanine motifs in orange.

4.5 Conclusions

Five unrestrained CG simulations is insufficient to establish the preferred interaction states of EphA1 and EphA2 with absolute confidence. However, the CG model does seem to suggest a preference for interaction via the glycine zipper motifs for both helices. The CG model provided a dimer similar to an observed experimental structure for EphA1, but for EphA2 the most stable interaction was one which has not been observed experimentally. The fact that an atomistic simulation of EphA2 interacting via both the glycine zipper and heptad motifs shows the helices seemingly in the process of dissociating may be explained by the complex relationship between a TM helix and its environment, with different lipid types having a significant effect on dimer stability. It also underlines the importance of validating predictions using a variety of methods. It would be illuminating to perform a larger number of atomistic simulations on both Eph dimer systems to obtain a more confident prediction of the preferred interaction modes of the Eph receptor helices.

Chapter 5

PMF Calculations of Eph Receptor Transmembrane Helices

5.1 Introduction

The dimerisation of helices in a membrane is a topic of current interest owing to its connection with signalling processes [209], as well as implications for membrane protein folding. Unrestrained CGMD simulations in Chapter 4 suggested that EphA2 can interact via both the glycine zipper and heptad motifs. If the rotation coupled activation mechanism does indeed describe the mechanism of EphA2 TM dimerisation, as suggested by Bocharov et al. [141], then these different interaction states may represent the active and inactive forms of the dimer. It would then be reasonable to expect the two interaction states to be similar in energy, with a barrier between the states higher than thermal energy but low enough to allow the receptor to transition between them without requiring a prohibitive energy input. As it is not feasible to quantify the full free energy landscape for the dimer from unrestrained simulations, umbrella sampling was used to calculate the PMF of the interaction between EphA2 (and EphA1) helices, with the separation of the helices used as the reaction coordinate. A fully sampled PMF profile is able to show the relative energies of the different states interacting via the different motifs, and so be able to test the above assumptions relating to the rotation-coupled activation mechanism. A similar PMF calculation was carried out on the TM helices of EphA1 to investigate the stability of the dimer with respect to EphA2, and the relative stability of the glycine zipper compared with the alanine motif.

To the best of my knowledge, no computational or experimental methods have yet been used to assess the stability of the Eph receptor TM dimers. However, a variety of compu-

tational techniques have been applied to the problem of calculating the PMF profile for the association of a variety of other TM helices. One such technique uses a mesoscopic model to calculate the PMF of simple proteins using dissipative particle dynamics (DPD) [119], which showed PMF depths of ∼-16 to -21 kJ/mol. In that example, the PMFs were dependent on the hydrophobic matching between the protein and bilayer. Atomistic MD simulations with umbrella sampling have been used to study the artificial peptide pVNVV [238], while atomistic simulations with the adaptive biasing force method [175, 176] have been used to study glycophorin A (GpA) [239] in a membrane mimetic. Both calculations were limited to a simulation length of 15 ns, which may not provide exhaustive sampling of all rotational degrees of freedom. CG models allow for much longer simulations, and the MARTINI model has been applied to calculate the PMF of a variety of systems including: GpA using MD [240]; GpA [241] and the ErbB receptors [242] using Monte Carlo (MC) simulations; and a variety of leukocyte integrins [243]. PMF depths and association free energies vary significantly depending on the method, environment and dimer studied: PMF depths for GpA vary from -25 kJ/mol for CGMC [241] (depending on lipid bilayer width) to -46 kJ/mol for atomistic ABF calculations [239]. The PMF depths for the leukocyte integrins vary from -32 to -53 kJ/mol depending on the helix sequence [243].

The results of such PMF calculations can be compared with results from experimental methods of measuring the association free energies of TM dimer systems. These can be determined using techniques such as sedimentation analytical ultracentrifugation (SE), fluorescence resonance energy transfer (FRET), thiol disulfide interchange (TDI), and the TOXCAT and related assays [244]. They have been applied to a variety of systems, including for example glycophorin A (GpA) [245, 246], the flu M2 tetramer [247, 248], and the erythropoietin dimer [249] - for a review see [250] or [251]. However, it is apparent that the measured association free energies are very dependent on environment; the same group using the same method (FRET) measured the association free energy of GpA varying from ∼-15 to -30 kJ/mol depending on the type of detergent used [252].

This dependence of the association free energy on the environment, seen in both simulation and experiment, makes comparisons between simulation and experiment difficult. This is exacerbated by the fact that the majority of experimental measurements take place in micelles. Helices in micelles have a higher degree of rotational entropy than in bilayers [253], and are subject to less lateral pressure, so the free energy of association of a system in micelles is likely to be significantly different from that measured in bilayers. Neverthe-

less, as GpA has been well characterised by both experiment and computation, it can be used to show that the two methods are at the least in qualitative agreement. Measures of the association free energy of GpA from experiment range from ∼15 kJ/mol in C_{11}DAO micelles [252] to ∼24 kJ/mol in C14 betaine micelles [254] and ∼38 kJ/mol in C_8E_5 micelles [255]. These compare reasonably well with the range of association free energies of GpA from computational methods, ranging from 25 [241] to 46 kJ/mol [239]. Probably the best-sampled estimate comes from Sengupta et al., who calculated a value of 38 kJ/mol [240] - or 47 kJ/mol if normalised to the appropriate standard state - in a DPPC bilayer using CGMD MARTINI simulations extended to 8 μs. This is slightly larger than the largest values measured in micelles, which is reasonable as it is likely that both the extra rotational entropy cost of dimerisation and lack of lateral pressure in micelles will result in a lower association free energy in micelles compared with bilayers.

It seems that computational methods can produce estimates of the free energy of association which are comparable with experimental values, given the high dependence of measurements on the environment used. Computational PMFs have also shown high dependence on bilayer width [241], helix type [243, 242] and mutations in the sequence [240]. Whether it is feasible to achieve the sampling necessary to be able to accurately estimate the energy differences between two different interaction states of the same dimer has yet to be established.

This chapter aims to establish:

1. the association free energy of dimerisation for EphA1 and EphA2 predicted by the CG model, and whether this estimate is comparable to experimental measurements and other computational estimates for similar systems;

2. the factors which influence the estimate of the association free energy; and

3. the difference in energy between the interactions via the heptad and glycine zipper motifs for EphA2.

5.2 Methods

5.2.1 Theory

As discussed in Chapter 2, the potential of mean force (PMF) $\mathcal{W}(\xi)$ along some chosen coordinate ξ can be written as

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right] \tag{5.1}$$

with ξ^* an arbitrary value on the chosen coordinate. $\langle \rho(\xi) \rangle$ is the average distribution function, defined as

$$\langle \rho(\xi) \rangle = \frac{\int \delta(\xi'(\mathbf{R}) - \xi) e^{-U(\mathbf{R})/k_B T} d\mathbf{R}}{\int e^{-U(\mathbf{R})/k_B T} d\mathbf{R}} \tag{5.2}$$

For the helix-helix PMF calculations in this chapter, $U(\mathbf{R})$ is the total energy of the simulated system. The reaction coordinate ξ is the separation between the centre of masses of the chosen motifs; this was chosen to ensure adequate sampling of the states interacting via the motifs. $\xi'(\mathbf{R})$ represents the separation of the centre of masses of the motifs in a given simulation. The integral $\int d\mathbf{R}$ is over all non-sampled degrees of freedom in the system.

5.2.2 Relationship between PMF and experimental results

From sedimentation and FRET experiments using micelles [256, 253], the apparent association constant measured is

$$K_{exp} = \frac{n_{Dimer}/V_{Aq}}{(n_{Monomer}/V_{Aq})^2} \tag{5.3}$$

where n_i are the number of moles of the ith species, and V_{Aq} the volume of the aqueous solution. To eliminate the dependence of K_{exp} on micelle concentration, the mole fraction equilibrium constant is usually reported, defined as

$$K_X = K_{exp} \times [Detergent] \tag{5.4}$$

$$= \frac{n_{Dimer}/n_{Total}}{(n_{Monomer}/n_{Total})^2}. \tag{5.5}$$

However, equation 5.4 contains an implicit dependence on the concentration of protein in the hydrophobic phase. It is conventional to take as a standard state 1 M of protein in hydrophobic phase. To introduce this explicitly consider the molar association constant

$$K_C = \frac{n_{Dimer}/V_{HP}}{(n_{Monomer}/V_{HP})^2} \tag{5.6}$$

where V_{HP} is the total volume of hydrophobic phase including protein and detergent molecules. The relationship between between K_X and K_C, assuming the system is in the dilute-ideal limit, can be shown to be [253]

$$K_C = K_X \times \mu_{Detergent} \tag{5.7}$$

for micelles, where $\mu_{Detergent}$ is the molar volume of hydrophobic tails. The standard state association free energy reported is then

$$\begin{aligned} \Delta G^{\circ}_{exp} &= -RT \ln K_C & (5.8) \\ &= -RT \ln K_X - RT \ln \mu_{Detergent}. & (5.9) \end{aligned}$$

The association constant as measured from the PMF is

$$K_{pmf} = \int_0^{\xi_{sep}} \exp\left(\frac{-\mathcal{W}(\xi)}{k_B T}\right) d\xi \tag{5.10}$$

where ξ_{sep} is the value of ξ between the dimeric and monomeric states, chosen to be the smallest value of ξ which is constant while the helices are in a monomeric state. K_{pmf} is not very sensitive to the choice of ξ_{sep}. Note that the integration in equation 5.10 is using Cartesian coordinates, as the motifs were restrained in both x and y dimensions, with the restraint varying along one axis only. Similar approaches have used polar coordinates, which would require the introduction of a Jacobian factor into equation 5.10.

Equation 5.10 contains an implicit dependence on the concentration of protein. To see this, consider K_{pmf} as a description of a two state reaction with a particular protein concentration:

$$K_{pmf} = \frac{[Dimer]/[Protein]}{([Monomer]/[Protein])^2}. \tag{5.11}$$

$[Dimer]$ is the concentration of protein dimers in the system; $[Monomer]$ is the concentration of protein monomers in the system; $[Protein]$ is the total concentration of protein molecules. As with the experimental results from micelles, it is conventional to define our standard state as 1 M of protein in hydrophobic phase. The standard state free energy change is then

$$\Delta G^{\circ}_{pmf} = -RT \ln K_s \tag{5.12}$$

with

$$K_s = \frac{[Dimer]/1M}{([Monomer]/1M)^2}. \tag{5.13}$$

The standard state free energy change can then be expressed as

$$\begin{aligned} \Delta G^{\circ}_{pmf} &= -RT \ln K_s & (5.14)\\ &= -RT \ln(K_{pmf}/[Protein]) & (5.15)\\ &= \Delta G_{pmf} + RT \ln[Protein] & (5.16) \end{aligned}$$

It may also be useful to define the depth of the PMF:

$$\Delta \mathcal{W}_{WD} = \mathcal{W}(\xi_{dimer}) - \mathcal{W}(\xi_{monomer}) \tag{5.17}$$

where $\mathcal{W}(\xi_{dimer})$ is the minimum value of the dimer portion of the PMF, and $\mathcal{W}(\xi_{monomer})$ is the value of the PMF when the helices are sufficiently separated that they no longer interact. $\mathcal{W}(\xi_{monomer})$ is chosen to be zero for these helix-helix PMFs.

5.2.3 Simulation Details

The starting positions for the PMF windows were the final structures from simulations in Chapter 4. The starting structure for the PMF with centre of mass of the glycine zipper motif restrained was taken from simulation 2 from the EphA2 simulations discussed in Chapter 4; this shows the helices interacting via the glycine zipper motif (i.e. in state G_1). The starting structure for the PMF calculation restraining the centre of mass of the heptad motif came from simulation 4 from the EphA2 simulations discussed in Chapter 4 (i.e. state H_1).

The centre of mass of the relevant motifs were restrained relative to each other with a harmonic restraint, with force constant of 1000 kJ/mol/nm^2. Restraints were applied in both x and y dimensions (x and y are orthogonal axes in the plane of the bilayer). The y restraint was set to zero for every simulation. The x restraint was set to the value of the reaction coordinate required for the window, varying from 0.4 nm to 2.7 nm in 0.1 nm intervals - snapshots from selected simulations are shown in Figure 5.1. The helices were pulled rapidly ($\sim$1-3 ns) to near the equilibrium point of the restraint, so the system required some time to subsequently equilibrate (see Section 5.4.1). The forces applied in the x direction by the restraints to the centre of mass of the helices were recorded. These were then unbiased using the Grossfield implementation of WHAM [195] (see Section 5.4.2 for further

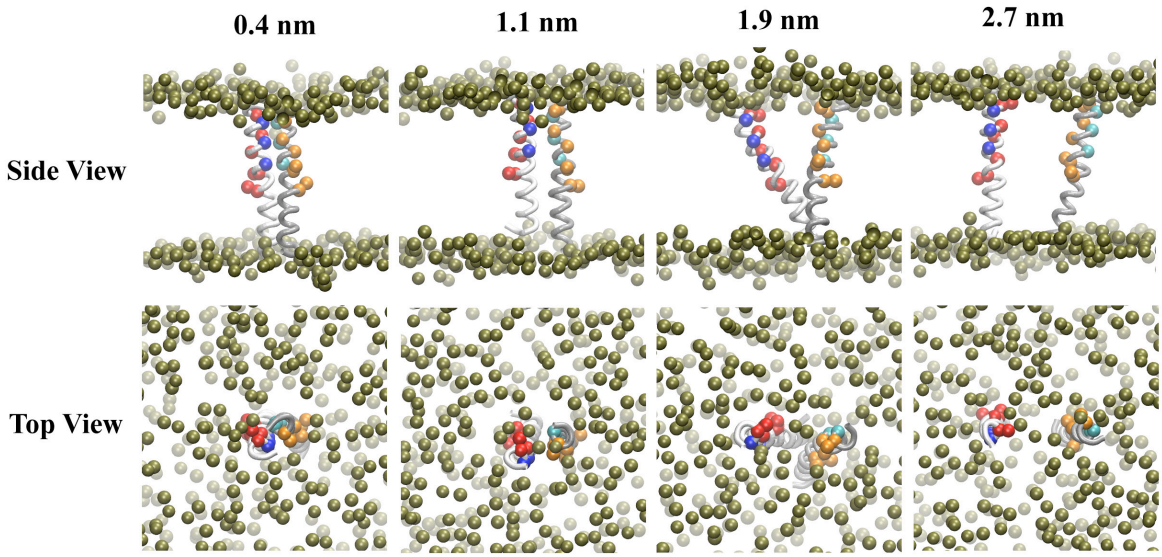


Figure 5.1: Snapshots from selected simulations which were used to calculate the PMF of EphA2, with the glycine zipper motif restrained. The equilibrium restraint separation is above the snapshots. The glycine zipper motifs are shown in blue and cyan, the heptad motifs in red and orange. Lipid headgroups are tan. Lipid tails, waters and non-motif sidechains are omitted for clarity.

discussion). The number of bins used was 200 with a tolerance of 0.0001.

All simulations in this chapter were performed using Gromacs version 4.5.3 [177]. The temperatures of the helix, lipids and solvent were each coupled separately using the Berendsen algorithm [153] at 323 K. The time step was 40 fs. The pressure was semi-isotropically coupled using the Berendsen algorithm at 1 bar with τ_P = 1 ps and a compressibility of 5 $\times$ 10^{-6} bar^{-1}.

5.2.4 PMF Calculations

All PMF calculations that were performed on the Eph receptor dimers are summarised in Table 5.1. A total of five PMF calculations were performed on EphA2. These were divided into 'ideal' helices, created using the 'build' function of Pymol with harmonic restraints applied between the i^{th} and $(i+4)^{th}$ Cα particle; and 'NMR' helices, which were produced by creating a coarse-grained model of the NMR structure of Bocharov et al. [141]. Restraints were applied only to those Cα pairs within 0.35 nm, not to every i^{th} and $(i+4)^{th}$ backbone pair of the NMR helices. These were therefore more flexible than the ideal helices - there are 27 such restraints for the 'ideal' helices, compared with 20 for the 'NMR' helices. PMF calculations were performed with restraints applied to either the centre of mass of the glycine zipper or heptad motif for each helix type. This was to assess whether adequate rotational

Table 5.1: Summary of PMF simulations. 'Model' refers to the secondary-structure restraints applied to the helices. 'Ideal' means the helices were restrained as ideal helices; 'NMR' means that the helix restraints were based on the NMR structure from Bocharov et al. [141]. See text for more details.

Helix	Model	Reaction coordinate	Simulation length per window (ns)
EphA2	Ideal	Gycine zipper	7500
EphA2	Ideal	Heptad	2500
EphA2	NMR	Gycine zipper	1500
EphA2	NMR	Heptad	1500
EphA2	NMR	Cα	1500
EphA1	Ideal	Gycine zipper	2500

sampling was provided by restraining just one of the motifs, or whether different conformations were sampled depending on where the restraint was applied. A PMF calculation was also performed using the separation of the centres of mass of all Cα particles for the EphA2 NMR helices as the reaction coordinate to further determine the effect of applying the restraints to the motifs. The PMF for the association of ideal EphA1 TM helices was also calculated, applying the restraints to the glycine zipper motif.

5.3 Results

5.3.1 EphA2 - Ideal Helix

Using the Glycine Zipper Motif as the Reaction Coordinate

The PMF for the association of ideal EphA2 TM helices, using the separation of the glycine zipper motifs as the reaction coordinate, is shown in Figure 5.2. This shows a PMF depth of $\sim$-55 kJ/mol, and what appear to be two minima at a separation of 0.5 nm (labelled A) and 1.2 nm (labelled B). Representative pictures of the helices at the two minima are shown in Figure 5.3. These pictures are the most representative frames from the simulation according to cluster analysis performed on the Cα particles of the helices [235]. The pictures indicate that the first minimum, at 0.5 nm separation, corresponds to the helices interacting via the glycine zipper motif with a shallow crossing angle. In contrast, the (shallow) minimum at 1.2 nm separation appears to correspond to the heptad motifs interacting in a right-handed dimer, with a larger crossing angle than the glycine zipper motif.

Analysis of the relative rotation angles of the helices can provide more information about the different states sampled in the simulations. Figure 5.4 shows the average relative rotation

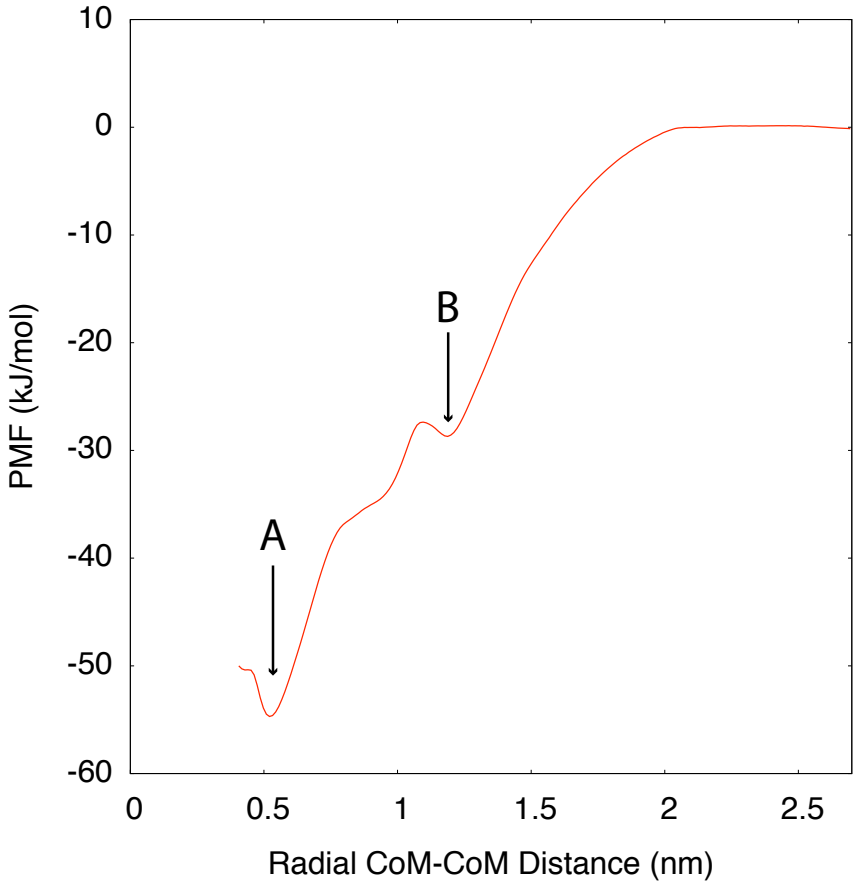


Figure 5.2: PMF of association of ideal EphA2 helices with the centre of mass of the glycine zipper motifs restrained. The PMF is calculated over the last 7 μs of simulation time. A marks the energy minimum when the motifs are restrained to a distance of 0.5 nm, B marks the minimum where the motifs are restrained to a distance of 1.2 nm.

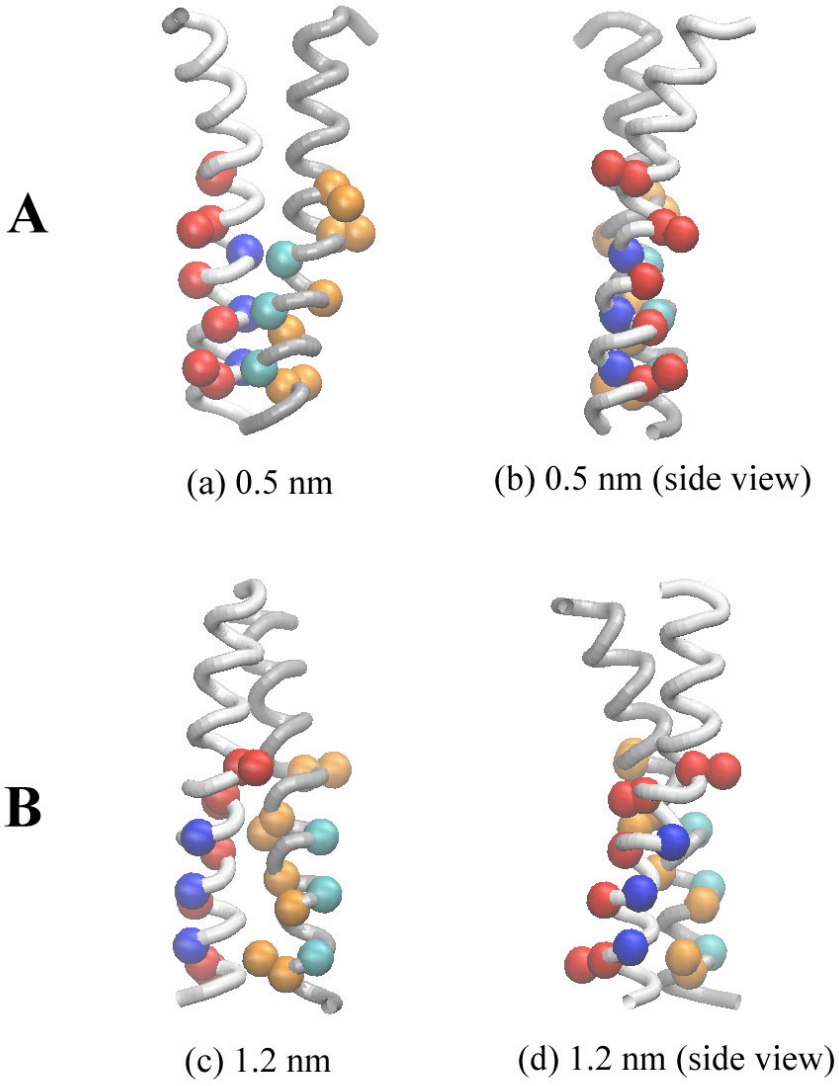


Figure 5.3: Snapshots from the windows corresponding to minima in the EphA2 PMF with the glycine zipper motifs restrained. The frame taken was the most representative according to cluster analysis [235]. A indicates the minimum at a motif separation of 0.5 nm, B indicates the minimum at a separation of 1.2 nm. Waters, lipids and non-motif side chain particles are omitted for clarity. The glycine zipper motif is shown in blue and cyan, the heptad motif in red and orange.

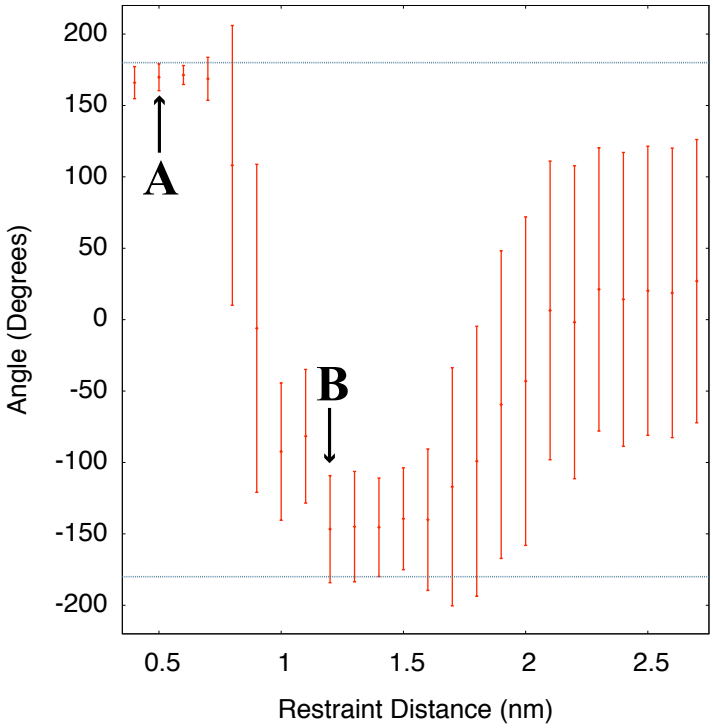


Figure 5.4: The rotation angle of the helices as a function of the window restraint. An angle of -180° indicates the heptad motifs interacting; +180° indicates the glycine motifs interacting. Error bars are standard deviations. Horizontal grey lines indicate the maximum and minimum values of 180° and -180°. The minima at a motif separation of 0.5 nm (A) and 1.2 nm (B) are marked.

angle across each window, where the angle is between the two vectors between the centre of mass of the glycine zipper and heptad motifs, in the plane of the bilayer. The angles are defined as in Chapter 4 such that an angle of +180° therefore corresponds to the glycine zipper motifs facing each other; -180° corresponds to the heptad motifs facing each other; 0° corresponds to opposite faces of the helices interacting. Figure 5.4 shows that the glycine zipper motifs sample only a small range of rotations when they are restrained to be within 0.4 - 0.7 nm of each other, suggesting that the helices interact tightly through the glycine zipper motif. The rotation angles suggest that the heptad motifs interact with each other when the glycine zipper separation is 1 - 1.6 nm. As the separation gets larger, the average rotation goes to zero, with large standard deviations, suggesting the helices are rotating independently of each other.

The helix rotation angles as a function of helix-helix centre of mass separation are shown in Figure 5.5. A similar range of rotation angles of -150 to -180° and +150 to +180° is seen for the glycine zipper and heptad interaction states. The two states also favour similar helix-helix separations, of ∼0.7 - 1.1 nm; however, the heptad interaction seems to occur over a slightly more diffuse range of helix-helix separations than the glycine zipper interaction

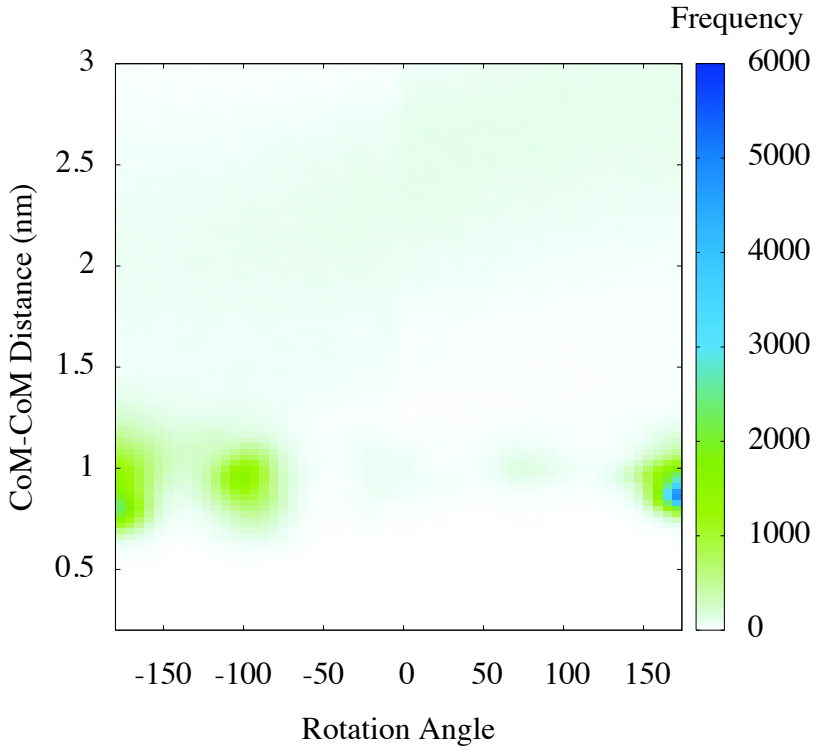


Figure 5.5: The rotation angle of the helices as a function of the helix centre of mass separation for the ideal EphA2 PMF with the reaction coordinate as the glycine zipper motif separation. An angle of -180° indicates the heptad motifs interacting; +180° indicates the glycine motifs interacting.

(∼0.7-1.2 nm compared with ∼0.8-1.1 nm). This suggests that the heptad interaction may be slightly less stable. Figure 5.5 also shows a third possible interaction state, with a rotation angle of ∼-100°, similar to state H_2 seen in the unrestrained simulations 4 and 5 of Chapter 4. As with the unrestrained simulations, the interactions via both the glycine zipper and heptad motifs produce a right-handed dimer.

Using the Heptad Motif as the Reaction Coordinate

The PMF of EphA2 with the separation of the heptad motifs as the reaction coordinate is shown in Figure 5.6. The PMF depth is ∼-47 kJ/mol, about 8 kJ/mol less than the PMF with the glycine zippers restrained. There are two significant wells, corresponding to a heptad separation of 0.5 nm (labelled C) and 0.9 nm (labelled D). The first minimum (C) represents the heptad motifs interacting, as shown in Figure 5.7, with a right-handed dimer similar to the minimum seen at 1.2 nm (B) separation in the PMF with the glycine zipper motifs restrained. The minimum at 0.9 nm (D) shows the glycine zipper motifs interacting, with a

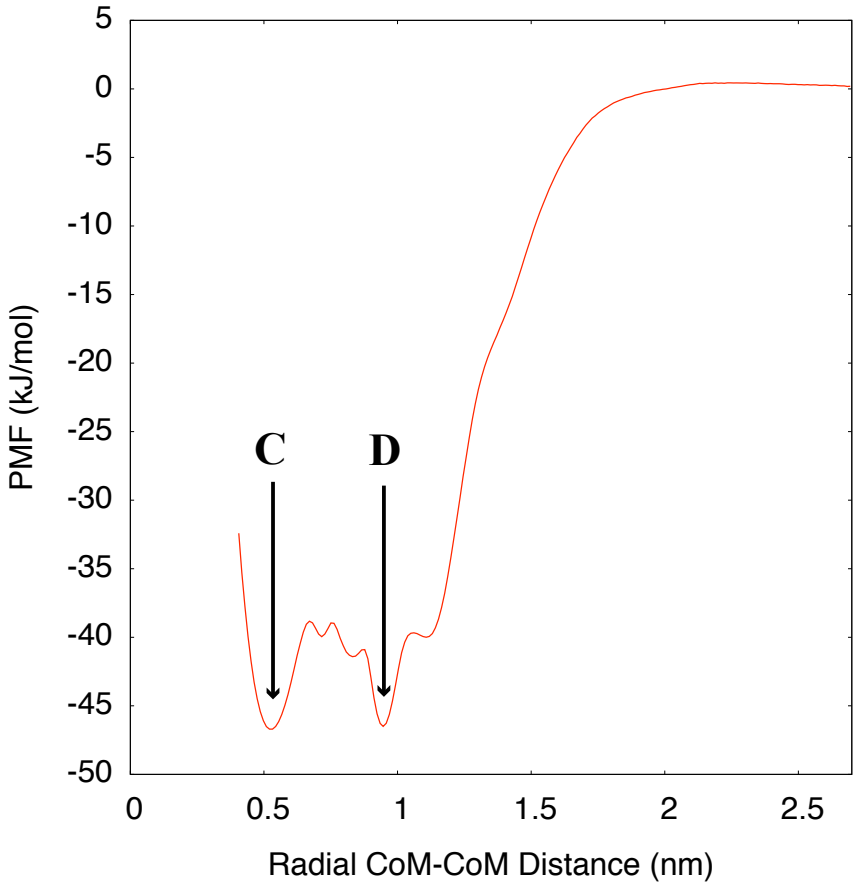


Figure 5.6: PMF of EphA2 with the helices restrained as ideal helices and the separation of the centres of mass of the heptad motifs as the reaction coordinate. The PMF is calculated over the last 2 μs of simulation time.

much shallower crossing angle than when the heptad motifs interact. Interestingly, the two minima seem to be at very similar energies, with a barrier of ~8 kJ/mol separating them.

The average rotation angles over each simulation show, as expected, the inverse situation to the glycine zipper-restrained simulations (Figure 5.8). For helix separations of 0.4 - 0.7 nm, the heptad motifs face each other (average angle of ~-170°), while the rotation angles sample a narrow range, shown by the small standard deviation. This suggests that the interaction state for these motif separations is quite stable. For heptad motif separations of 0.9 - 1.5 nm the glycine zipper motifs face each other (average angle of ~-170°), still sampling a relatively narrow range of rotation angles. The interaction state via the glycine zipper motifs therefore shows comparable flexibility to the interaction via the heptad motifs. The rotation angles as a function of helix-helix separation, shown in Figure 5.9, also show well-defined rotation angles for both interaction states. It further suggests that the heptad interaction occurs when the helices are on average slightly closer to each other (0.6 to 0.9 nm) than the glycine zipper interaction (occurring at a helix separation of 0.8 to 1.1 nm). This could be interpreted as an

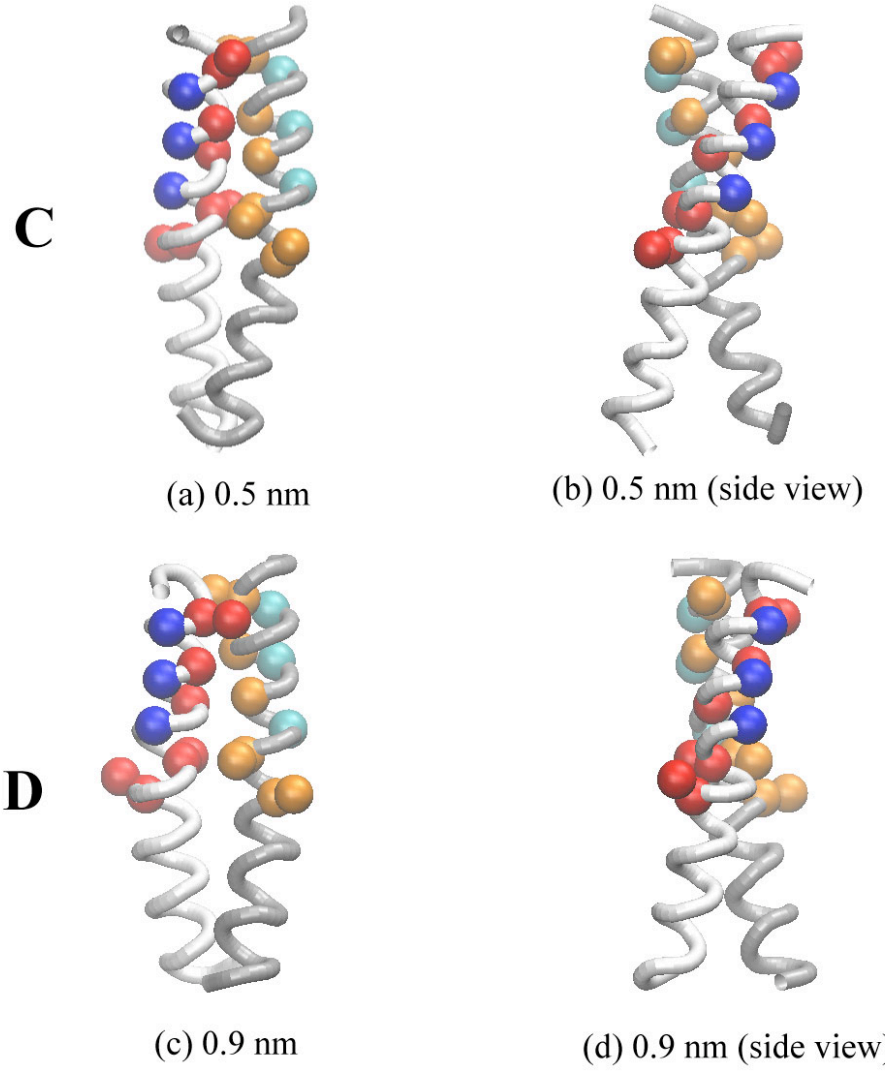


Figure 5.7: Snapshots from the windows corresponding to the minima of the EphA2 PMF calculation with the heptad motifs restrained. The frame taken was the most representative according to cluster analysis [235]. Waters, lipids and non-motif side chain particles are omitted for clarity. The glycine zipper motif is shown in blue and cyan, the heptad motif in red and orange.

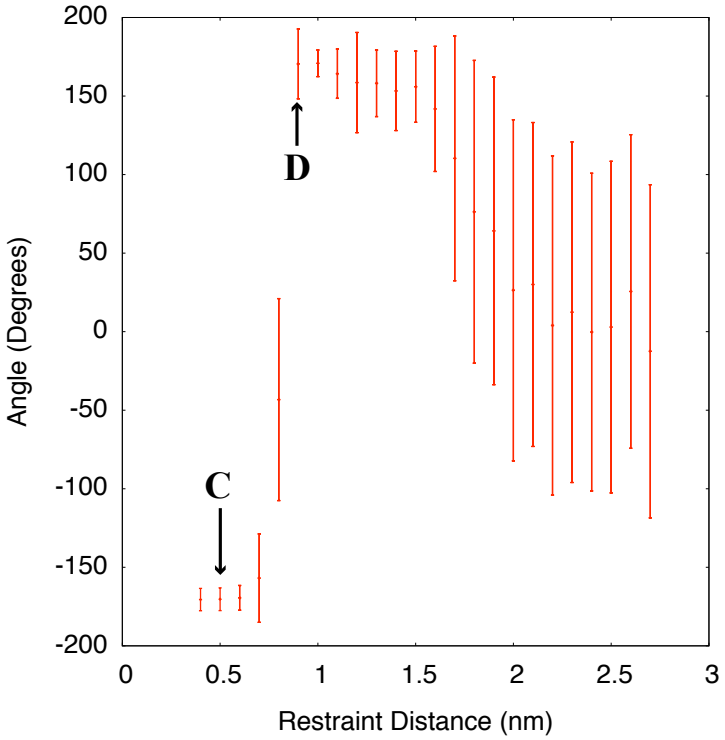


Figure 5.8: The rotation angle of the helices as a function of the window restraint for the PMF with the separation of the heptad motifs as the reaction coordinate. An angle of -180° indicates the heptad motifs interacting; +180° indicates the glycine motifs interacting.

effect of the restraints; however, the fact that the dimers in state H_1 were closer than in state G_1 in the unrestrained simulations in Chapter 4 suggests that the heptad interaction does indeed result in helices packing closer than the glycine zipper interaction. Figure 5.9 also shows no evidence for the third interaction state (H_2) at an angle of ∼100°, seen in Figure 5.5. This may be because of a different starting structure, or because the simulations are not long enough to capture transient interaction states.

5.3.2 EphA2 - NMR structure

PMF profiles were also calculated for the EphA2 TM helices derived from the NMR structure [140]. These helices were generated by applying the CG protocol directly to the structure; this produces helices which are more flexible than the 'ideal' helices, as harmonic restraints were applied to Cα particles within a cutoff distance of 0.35 nm, rather than to all $(i, i+4)$ pairs. The resulting PMF profiles are shown in Figure 5.10. These profiles are calculated from the last 1000 ns of a total window time of 1500 ns. This is unlikely to be sufficient to achieve full convergence but is likely to give a reasonable guide to which are the low energy states along the reaction coordinate (see discussion in Section 5.4.3 on sampling).

The PMF with the separation of the glycine zipper motifs as the reaction coordinate (the

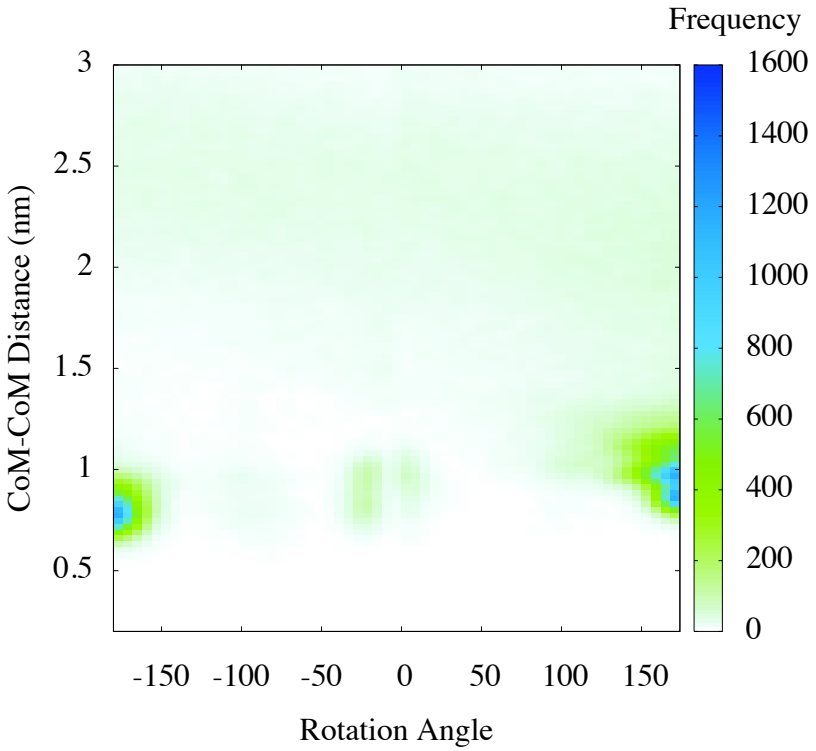


Figure 5.9: The rotation angle of the helices as a function of their separation for the PMF with the separation of the centres of mass of the heptad motifs as the reaction coordinate. An angle of -180° indicates the heptad motifs interacting; +180° indicates the glycine motifs interacting. The distribution shows roughly equal distributions between the heptad and glycine zipper motifs.

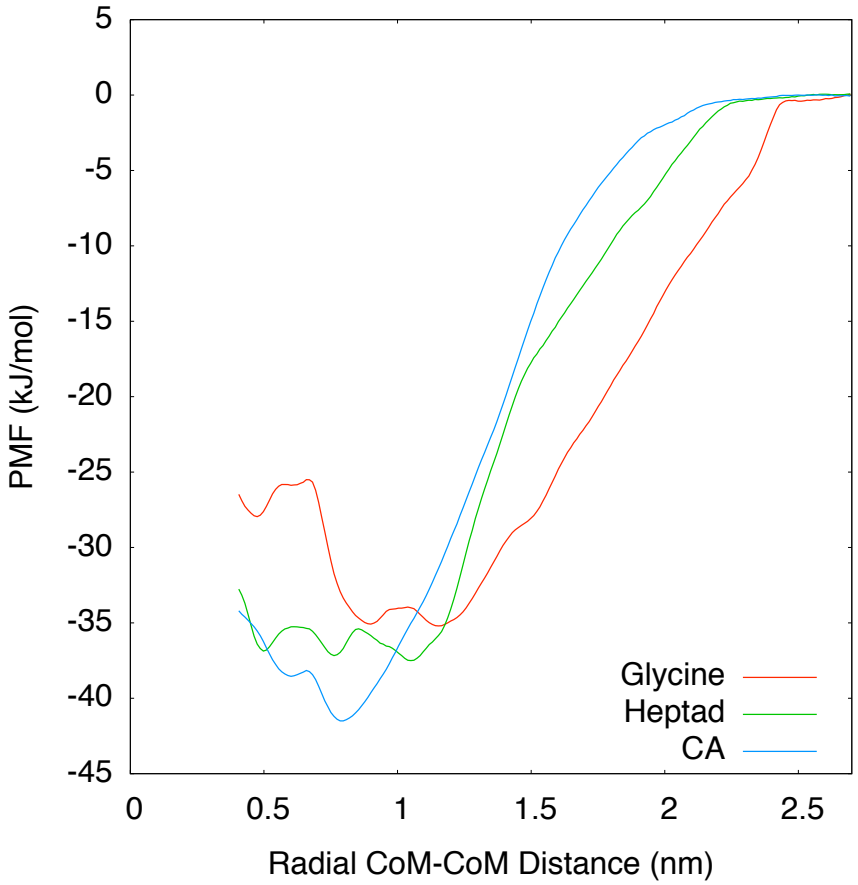


Figure 5.10: PMF profiles for EphA2 helices derived from the NMR structure. Profiles from calculations with restraints applied to centre of mass of the glycine zipper (red), heptad (green) and the central CA particles (blue). As each window was only 1.5 μs, these profiles are not likely to be fully converged.

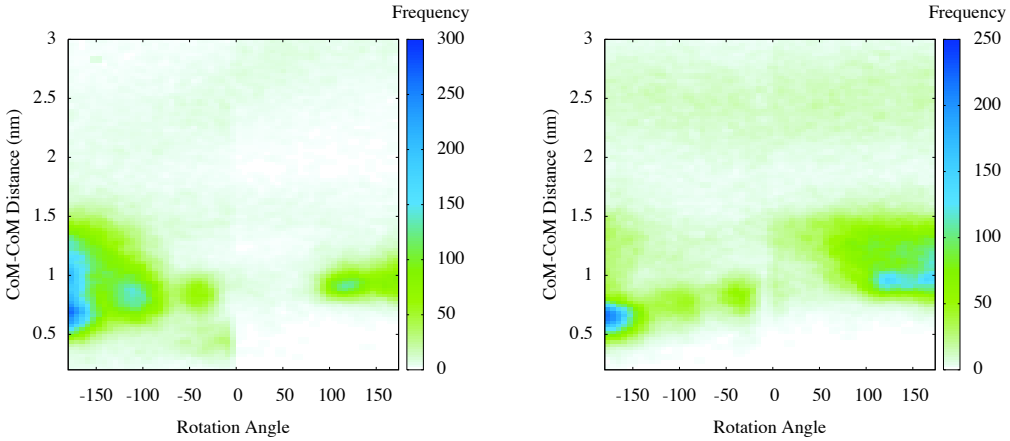
red line in Figure 5.10) has two minima at a motif separation of 0.9 and 1.2 nm. These are significantly lower in energy (∼10 kJ/mol) than states where the glycine zipper motifs are 0.4-0.8 nm apart. Representative snapshots from the simulations (Figure 5.13) show that these states involve the heptad motifs interacting (1.2 nm) or the helices interacting with different faces (0.9 nm). The average rotation angle from the simulations (Figure 5.12 (a)) shows that the windows with the glycine zipper motifs restrained between 0.4 to 0.6 nm are the only simulations in which these motifs face each other; similarly the rotation angle as a function of helix separation shows the heptad motifs face each other far more frequently than the glycine zipper motifs (Figure 5.11 (a)).

The PMF with the heptad motifs restrained (green line in Figure 5.10) shows three minima at similar energy levels. These are at a heptad motif separation of 0.5, 0.7 and 1.1 nm. The snapshots from these simulations suggest that the first two minima correspond to the heptad motifs interacting, while the third corresponds to the glycine zipper motif interacting (see Figure 5.14). The rotation angle suggests that the heptad motifs largely interact when the helix separation is 0.5 to 0.7 nm, while the glycine zipper motifs face each other when the helices are further apart (Figure 5.11 (b)). However, the structure with the heptad motifs facing each other seems to sample a narrower range of rotations than when the glycine zipper motifs face each other, suggesting that the former may be a more stable state - consistent with the PMF profile with the glycine zippers restrained. This may just be a result of the position of the restraints keeping the motifs close together, although the fact that the reverse pattern is not observed in the glycine-restrained PMF suggests otherwise.

The PMF with the Cα particles restrained (blue line in Figure 5.10) shows two minima at a separation of 0.6 and 0.8 nm. The first of these corresponds to the heptad motifs interacting; the second corresponds to a more complex structure where the glycine zipper motif of one helix is interacting with the heptad motif of the other helix (Figure 5.15). The rotation angles for these simulations show that the interactions with the heptad motifs are much better sampled than interactions via the glycine zipper (Figure 5.11 (c)).

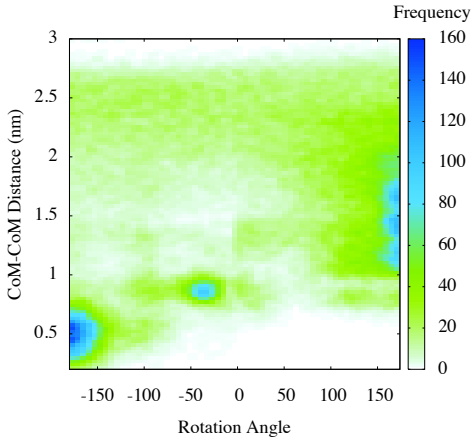
5.3.3 EphA1 - Ideal Helix

The PMF of the association of ideal EphA1 helices is shown in Figure 5.16, with a PMF depth of -45 kJ/mol, comparable to the EphA2 PMFs. It shows two significant minima, at a glycine zipper separation of 0.5 nm (labelled E) and 1.2 nm (labelled F). The first minimum (E) corresponds to the glycine zipper motifs interacting (see Figure 5.17); the second (F)



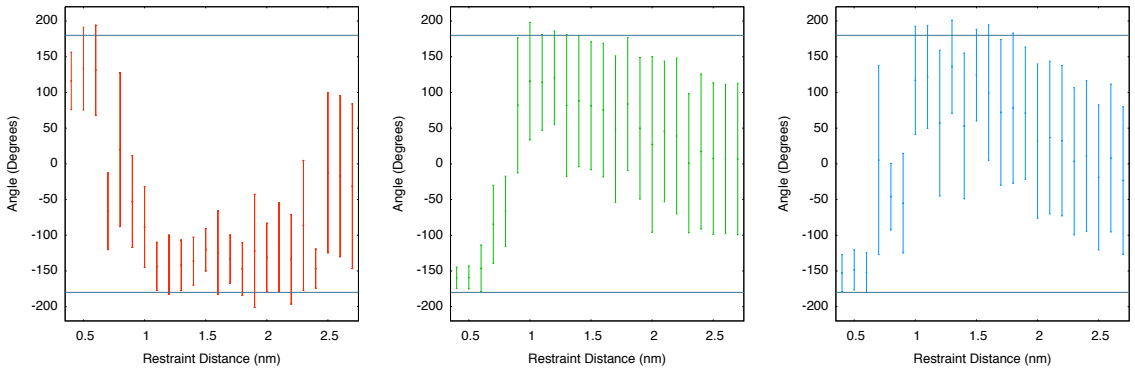
(a) Glycine

(b) Heptad



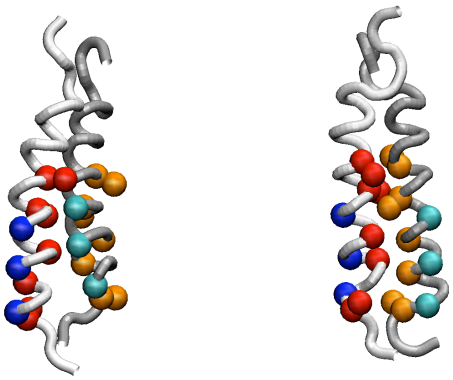
(c) CA

Figure 5.11: Rotation angle of EphA2 from NMR-derived helix simulations as a function of helix-helix separation. a) Simulations with centre of mass of the glycine zipper motifs restrained. b) Simulations with the centre of mass of the heptad motifs restrained. c) Simulations with the centre of mass of the Cα particles restrained.



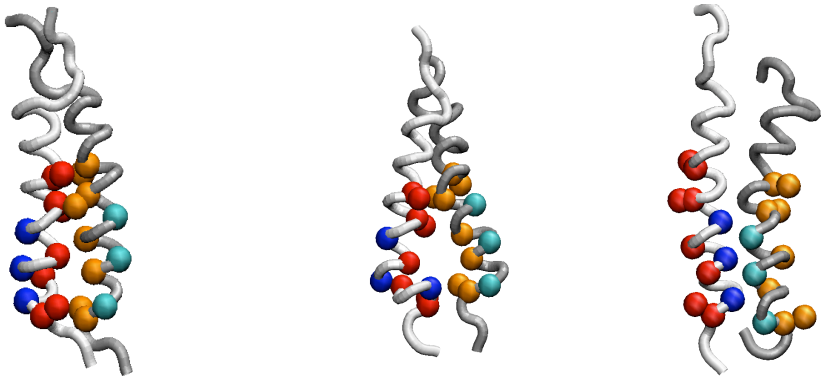
(a) Glycine (b) Heptad (c) Cα

Figure 5.12: Average rotation angle of EphA2 from NMR-derived helix simulations. a) Simulations with separation of the centres of mass of the glycine zipper motifs as the reaction coordinate. b) Simulations with the separation of the centres of mass of the heptad motifs as the reaction coordinate. c) Simulations with the separation of the centres of mass of the Cα particles as the reaction coordinate.



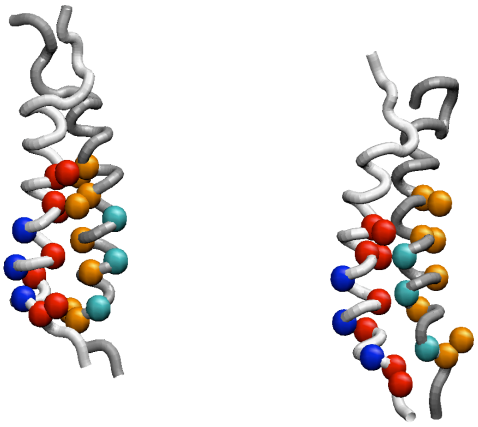
(a) 0.9 nm (b) 1.2 nm

Figure 5.13: Snapshots of EphA2 from NMR-derived helix simulations with separation of the glycine zipper as the reaction coordinate. Frames are the most representative from the simulation as indicated by cluster analysis [235]. Waters, lipids and non-motif sidechain particles are omitted for clarity. Heptad motifs are coloured red and orange; glycine zipper motifs are coloured blue and cyan.



(a) 0.5 nm (b) 0.7 nm (c) 1.1 nm

Figure 5.14: Snapshots of EphA2 from NMR-derived helix simulations with heptad separation as the reaction coordinate. Frames are the most representative from the simulation as indicated by cluster analysis [235]. Waters, lipids and non-motif sidechain particles are omitted for clarity. Heptad motifs are coloured red and orange; glycine zipper motifs are coloured blue and cyan.



(a) 0.6 nm (b) 0.8 nm

Figure 5.15: Snapshots of EphA2 from NMR-derived helix simulations with Cα separation as the reaction coordinate. Frames are the most representative from the simulation as indicated by cluster analysis [235]. Waters, lipids and non-motif sidechain particles are omitted for clarity. Heptad motifs are coloured red and orange; glycine zipper motifs are coloured blue and cyan.

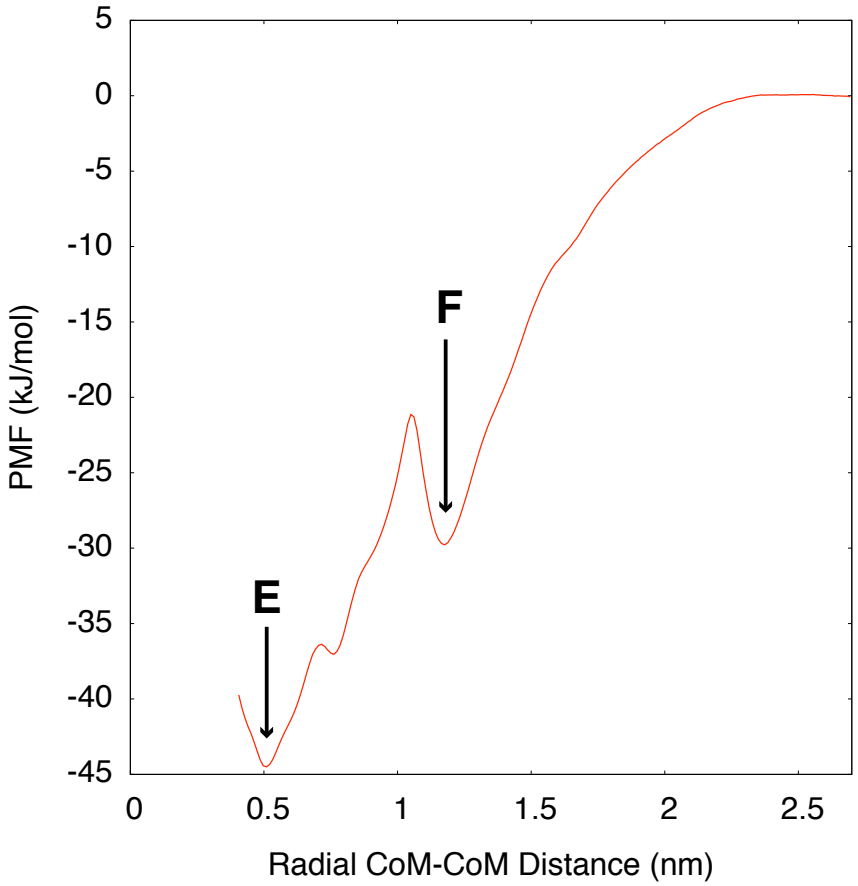


Figure 5.16: PMF of EphA1 with the helices restrained as ideal helices and the PMF restraint applied to the centre of mass of the glycine zipper motifs. The PMF is calculated over the last 2 μs of simulation time.

corresponds to a state where the helices interacting via neither the glycine zipper nor alanine motifs. There is a small crossing angle for both states.

As in Chapter 4, the rotation angle for EphA1 is the angle between the vectors defined by the centre of masses of the glycine zipper and alanine motifs in the plane of the bilayer. The angle is positive if the glycine zipper motifs are closer together than the alanine motifs; otherwise it is negative. The average rotation angles in the simulations (Figure 5.18) and the rotation angles as a function of helix-helix separation (Figure 5.19) show that a significantly larger range of rotation angles are sampled in the EphA1 simulations than in any of the EphA2 simulations. This suggests that the EphA1 structure is more flexible, and possibly less stable, than the EphA2 structures.

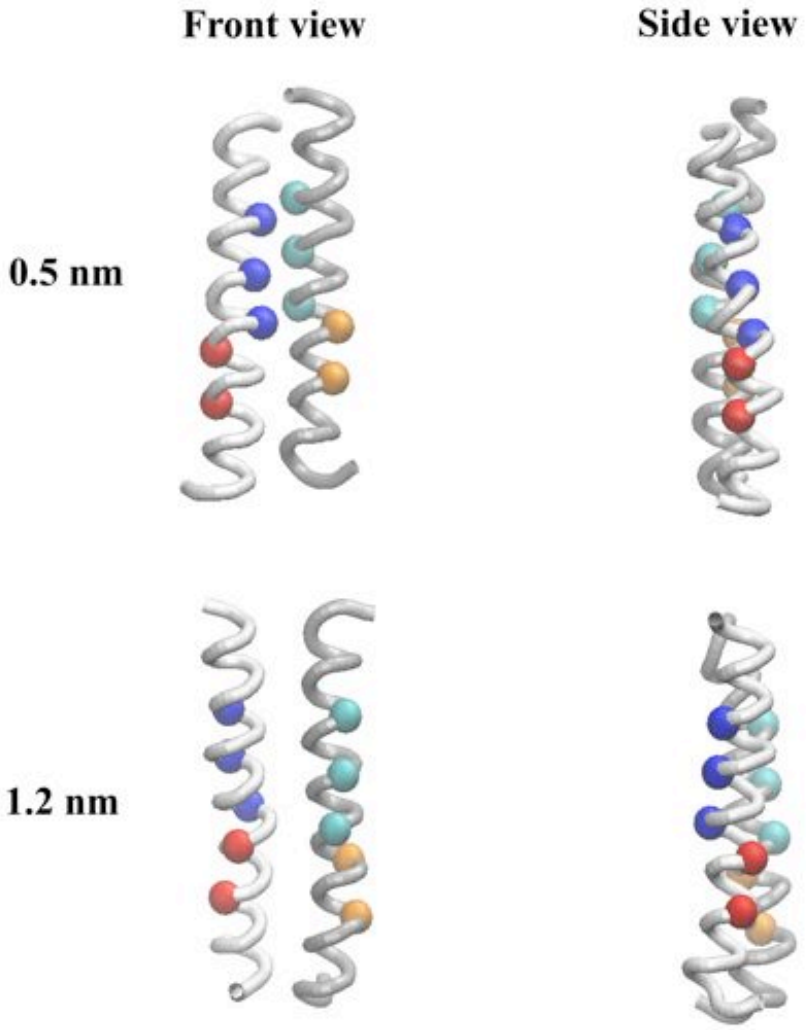


Figure 5.17: Snapshots from the selected windows of the EphA1 PMF calculation. The frame taken was the most representative according to cluster analysis [235]. Waters, lipids and non-motif side chain particles are omitted for clarity. The glycine zipper motif is shown in blue and cyan, the alanine motif in red and orange.

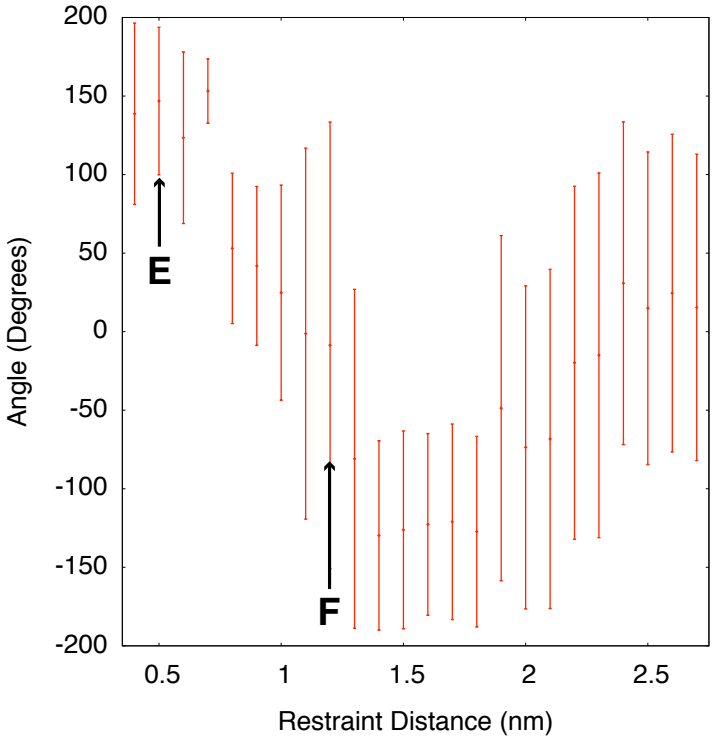


Figure 5.18: The rotation angle of the helices as a function of their separation for the EphA1 simulations. An angle of -180° indicates the alanine motifs interacting; +180° indicates the glycine motifs interacting.

5.4 Sensitivity Analysis

5.4.1 Calculating Equilibration Time

Each window starts with the same structure but applies different restraints to the motifs. In each simulation the helices therefore move rapidly from their starting position to the separation corresponding to the centre of the restraint, so the simulations required some equilibration time for the lipids and helices to adjust. The PMF of EphA2 with the separation of the glycine zipper motif as the reaction coordinate was chosen to investigate the equilibration time, as it has the longest simulations. The length of equilibration time was estimated by splitting the simulations into bins of varying lengths. ΔG_{pmf} was then calculated for all bin (time) widths; these are shown in Figure 5.20, which indicates that ΔG_{pmf} is estimated consistently for bin widths of 250 - 2500 ns. A plot of ΔG_{pmf} as a function of simulation time, for bin widths of 250, 500 and 1500 ns, shows that the estimate of ΔG_{pmf} for the first 250 ns is not significantly different than that calculated over different 250 ns periods of the simulation (see Figure 5.21), or from estimates using other bin sizes. This suggests that the equilibration time for this particular PMF is significantly less than 250 ns. This can be further confirmed by visual inspection of the forces on the helices applied as a result of

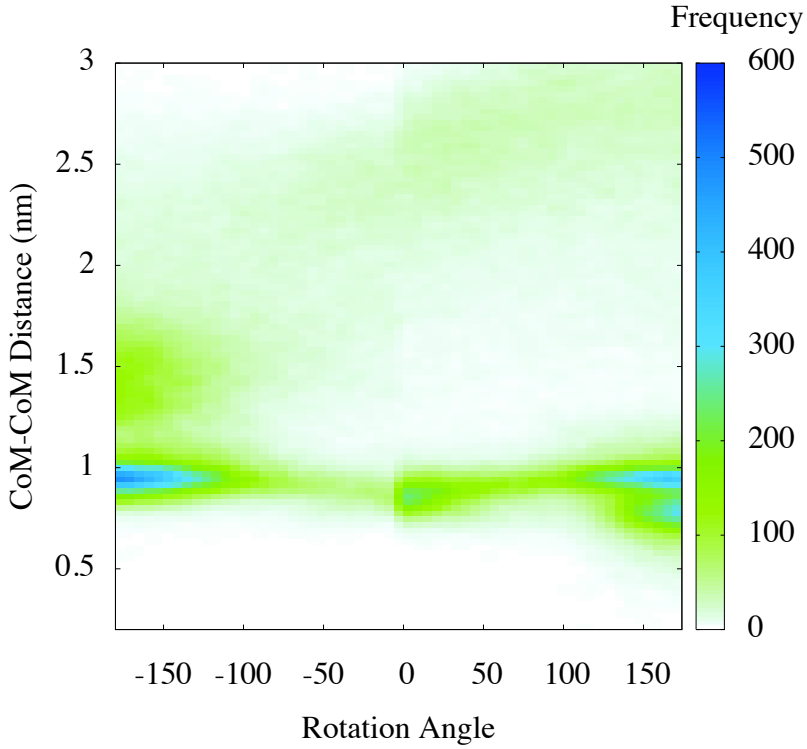


Figure 5.19: The rotation angle of the helices as a function of their separation for the EphA1 simulations. An angle of -180° indicates the alanine motifs interacting; +180° indicates the glycine motifs interacting.

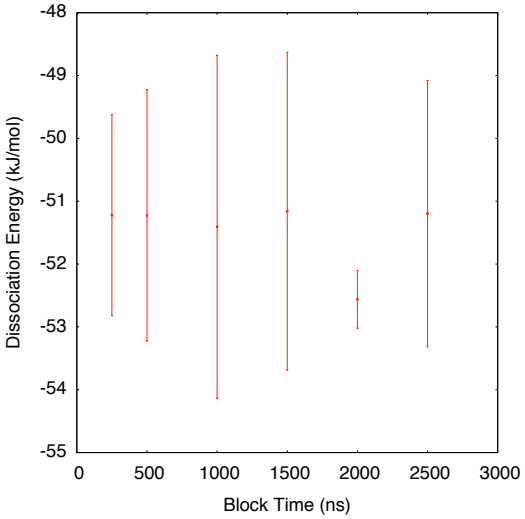


Figure 5.20: Values of ΔG_{pmf} calculated over a range of bin widths, for the PMF of EphA2 with the glycine zipper motif restrained. Error bars represent statistical errors. Bin widths of 250 - 2500 ns all provide very similar estimates of ΔG_{pmf}.

the restraint, shown for selected windows in Figure 5.22. This shows that the forces reach a consistent pattern on a timescale well below 500 ns. To allow a margin for error, the first 500 ns of data were discarded from all simulations, on the assumption that the equilibration time will be similar for all PMF calculations presented in this chapter. Visual inspection of the forces for simulation windows in other calculations confirmed that this assumption is reasonable (data not shown).

5.4.2 Implementing WHAM

Alan Grossfield's implementation of WHAM [195] has been widely used (for examples see [257, 258]). The Gromacs developers have recently introduced an alternative implementation [259]. To investigate the effects of the algorithm used, a PMF profile was produced using both methods over the time period 500-2500 ns for both the glycine zipper and heptad restrained EphA2 simulations. The results are shown in Figure 5.23, which shows that both algorithms produce profiles with similar shapes, with minima at the same values of the reaction coordinate. However, the depths are rather different, with significantly different values of ΔG_{pmf} for the profiles: -44.3 kJ/mol for the Grossfield algorithm compared with -56.2 kJ/mol for the Gromacs algorithm, for the heptad-restrained simulations; and -48.7 compared with -63.9 kJ/mol for the glycine zipper-restrained simulations. The source of this difference is unknown, but the fact that two different implementations of the same algorithm produce two significantly different results is a cause for concern. The remaining PMF profiles in this chapter have been calculated using the Grossfield implementations of

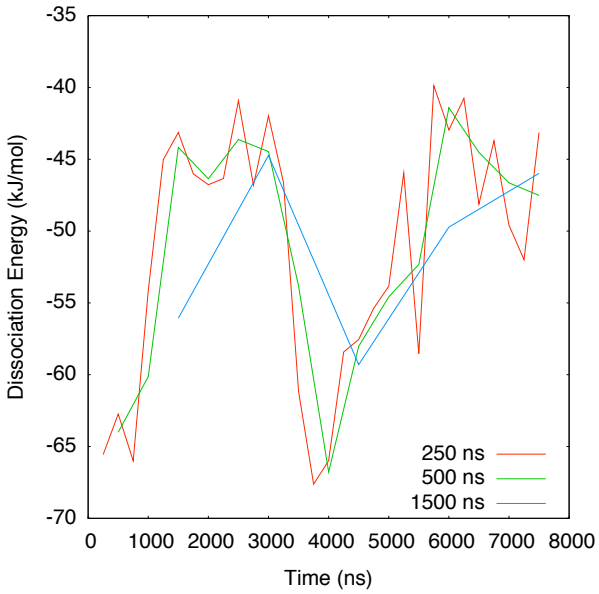


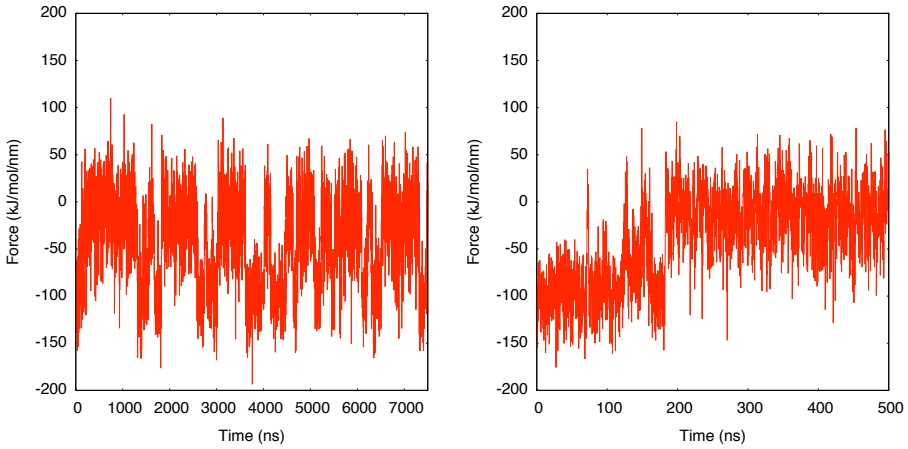
Figure 5.21: Values of ΔG_{pmf} as a function of simulation time for the PMF of EphA2 with the glycine zipper motif restrained. Values of ΔG_{pmf} when calculated over the first 250 ns are similar to those calculated over subsequent periods of the simulation, suggesting that the equilibration time is significantly less than 250 ns.

WHAM, which has been more widely used and so subject to more wide-ranging scrutiny.

5.4.3 Convergence

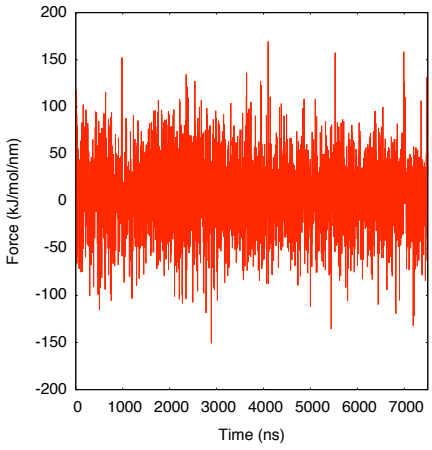
To draw any significant conclusions from the PMF profiles it is necessary to establish the level of sampling achieved in the simulations. To be confident that adequate sampling has been achieved, the simulation time should be much longer than the correlation time. In other words, the simulations should be long enough that multiple independent profiles can be calculated from the same simulation. To test this, the PMF profile was calculated for the glycine zipper-restrained ideal EphA2 simulations over consecutive 2000 ns time bins, from 1500-3500, 3500-5500 and 5500-7500 ns. The resulting profiles are shown in Figure 5.24 (a). These profiles show significant differences, with the PMF depth of the middle profile some 10 kJ/mol larger than the others, and the first profile showing a significant minimum at a motif separation of $\sim$1 nm, whereas the others show a point of inflection. These differences suggest that the correlation time is greater than 2000 ns.

Another test is to consider the effect of extending the simulation time on the profile. The PMF profile for the glycine zipper-restrained ideal EphA2 simulations calculated over increasing time periods - 500-2500, 500-4500, 500-6500 and 500-7500 ns - is shown in Figure 5.24 (b). These show that the shape of the profile has broadly been established by 2500 ns of simulation time, but there are significant ($\sim$5 kJ/mol) changes in the parts of the profiles as

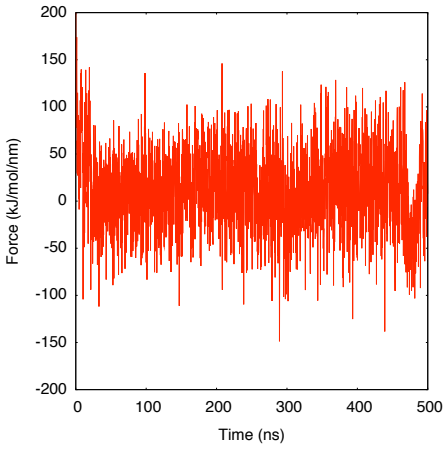


(a) 0.4 nm

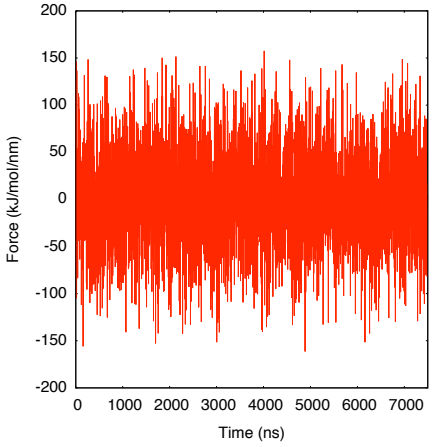
(b) 0.4 nm (first 500 ns)



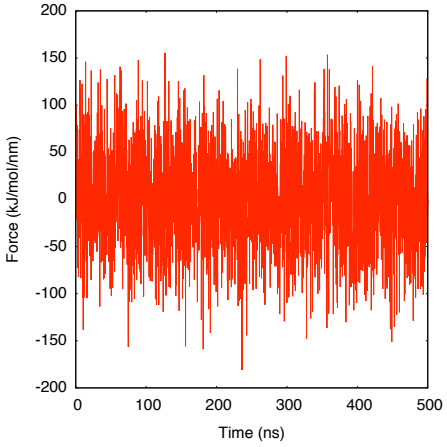
(c) 1.2 nm



(d) 1.2 nm (first 500 ns)

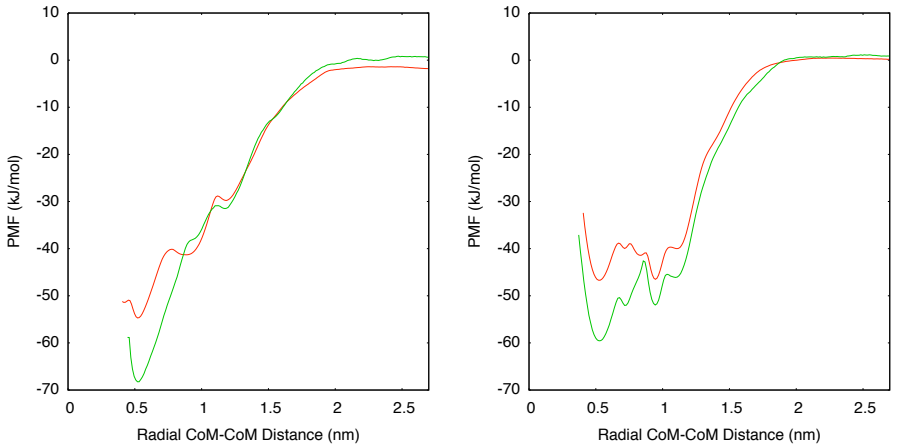


(e) 2.5 nm



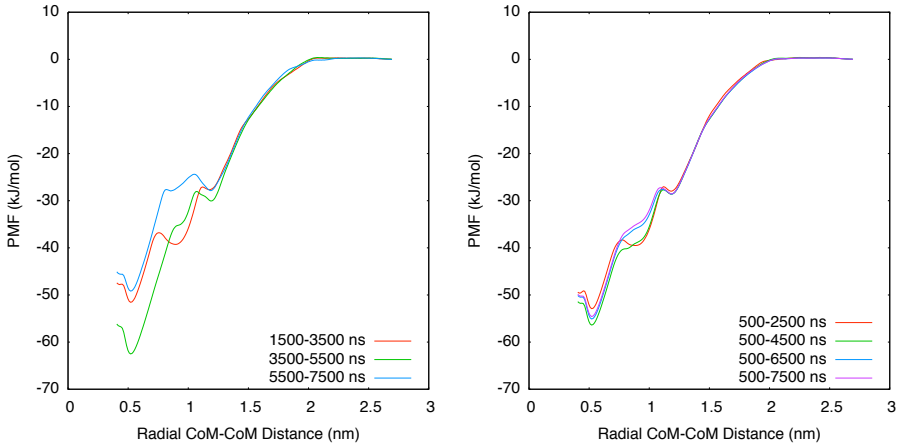
(f) 2.5 nm (first 500 ns)

Figure 5.22: Plots of the force acting on the restrained centre of mass of the glycine zipper motifs as a function of simulation time for selected windows for the PMF of the ideal EphA2 helices. These suggest that each window has an equilibration time much less than 500 ns.



(a) Glycine (b) Heptad

Figure 5.23: PMF profiles calculated using different implementations of WHAM, calculated for the period 500-2500 ns of the simulations. a) Glycine zipper motif restrained. b) Heptad motif restrained. Red: Grossfield algorithm [195]. Green: Gromacs implementation [259].



(a) (b)

Figure 5.24: PMFs of EphA2 with glycine zipper motifs restrained. (a) Profile calculated over consecutive 2000 ns periods of simulation time. (b) Profile calculated over gradually extended periods of simulation time.

the time over which they are calculated is extended from 2000 to 4000 ns. However, there are only very minor differences when the time is increased from 6000 to 7000 ns. This suggests that the profile may be converged after 7500 of simulation time, but is not conclusive, as the 'extra' simulation time is a small proportion of the total. It is possible that doubling the simulation time would allow the helices to sample new conformations and produce a profile which is significantly different.

Despite these caveats, some useful information may still be obtained from these profiles. As has been discussed in section 5.4.2, the calculated association free energy of the EphA2 dimer is the same, to within 1 kJ/mol, whether calculated over a simulation time of 250 ns or 2500 ns. This suggests that the association free energies of the dimer system can be reliably estimated after an (equilibrated) simulation time of $\sim$1000 ns. It is also worth noting that all profiles in Figure 5.24 show minima at the same two values of the reaction coordinate (0.5 and 1.2 nm), suggesting that useful information can be available in profiles which have not fully converged. However, reliably identifying what is useful and what is transient requires simulations significantly longer than the correlation time.

The PMF profiles in this chapter suggest that a simulation time of at least 7500 ns is needed to obtain a converged profile, and possibly even longer. Combined with different results being obtained from different implementations of the WHAM algorithm (see section 5.4.2 for more details), and different states being favoured when the flexibility of the helices is changed, it suggests that there are significant sources of error in these estimates beyond the statistical errors in association free energies that have been quoted. The profiles in this chapter should therefore be interpreted as at best a semi-quantitative picture of the energy landscape of these Eph receptor helices.

5.4.4 Effect of Choice of Reaction Coordinate

It is reasonable to expect a fully sampled PMF for EphA2 to show similar energetics for equivalent states, whether the reaction coordinate is defined through the glycine zipper or heptad motif separation. In other words, the value of the PMF when the helices interact via the glycine zipper motif should be similar, although the value of the reaction coordinate where this occurs would be expected to differ. In fact the difference in energy between equivalent states appears to be significantly different in the two profiles (see Figures 5.2 and 5.6); with the glycine zipper motifs restrained, the interaction via the glycine zippers seems to be $\sim$25 kJ/mol more favourable than interaction via the heptad motif, whereas the PMF

with the heptad motifs restrained suggests that the two states are at a similar energy level. Insufficient sampling in the latter PMF is likely to be a contributory factor - the discussion in section 5.4.3 notes that 2500 ns is insufficient for the heptad-restrained simulation to be fully sampled. However, a discrepancy of 25 kJ/mol is large, and extending the glycine-restrained simulations beyond 2500 ns changed the profile by only about 5 kJ/mol. This suggests that extending the heptad-restrained simulations may not fully resolve the discrepancy. It is also possible that restraining the centre of mass of the motifs in both x and y dimensions may limit the ability of the helices to rotate around each other fully, although analysis of the rotation angles from the EphA1 simulation suggests that in fact the helices are free to rotate with these types of restraint applied (Figure 5.19). Given that neither explanation is fully satisfactory for the difference in the two profiles, a two dimensional PMF calculation, with both the helix relative separations and rotations restrained, may be needed to confidently establish the full free energy landscape of the system.

Significant differences are observed in the behaviour of the dimer systems depending on the flexibility of the helices. Simulations of the NMR helices suggested a significant preference for the heptad motifs to interact, with little evidence of the helices interacting via the glycine zipper motifs. However, some snapshots suggested that the helices interact with different faces, and that the helices twist around each other in complex patterns (see Figures 5.13, 5.14 and 5.15). In contrast, the ideal helix simulations show helices primarily interacting via the glycine zipper or heptad motifs, with the glycine zipper motif favoured (according to the glycine-restrained PMF) or the two states at similar energy (according to the heptad-restrained PMF), although the necessity of interpreting the PMFs as a semi-quantitative description of the energy landscape should be emphasised.

Given the dependence of the behaviour of the dimer on the flexibility of the helices, it is necessary to identify which is likely to be a better model of the helices *in vivo*. Given the high energy penalty for unfolding a helix in the hydrophobic environment of the membrane interior [180], it seems likely that the TM regions of the Eph receptors are fully folded into helices. The flexibility of the NMR helices probably reflects a problem with the coarse-graining process - that not all the backbone particle pairs were within the cutoff to apply the harmonic restraints designed to mimic secondary-structure hydrogen bonds - rather than a genuine flexibility seen *in vivo*. This underlines the necessity of fully understanding the simplifications and assumptions underlying a computational model of a biological process. Further discussions of the results will therefore concentrate on the PMF profiles obtained

from the ideal helices.

5.5 Discussion

5.5.1 Comparison with Similar Systems

Table 5.2 shows the association free energies calculated from the PMFs. The ideal EphA1 and EphA2 PMFs show association free energies (in the standard state, i.e. ΔG°_{pmf}) between 52 and 62 kJ/mol. I am not aware of any experimental measures of the association free energy of the Eph receptor helices. However, these estimates can be compared with the estimates of the association free energy of the ErbB receptor TM helices, which are also from the RTK family, as calculated from CGMC simulations. These estimates varied from 26 kJ/mol for ErbB1 to 38 kJ/mol for ErbB4 [242]. At first glance these look considerably smaller than the estimates for EphA2 and EphA1; however, the same method (CGMC) produced an estimate for the association free energy of GpA in DPPC of 33 kJ/mol, which is considerably less than the estimate of 47 kJ/mol [1] from Sengupta and Marrink [240], also using MARTINI. This suggests that something in the CGMC method may provide lower estimates for the association free energy than CGMD methods (both were using the MARTINI CG model). The association free energies for EphA1 and EphA2 are comparable with the value of ΔG°_{pmf} of 47 kJ/mol for GpA from CGMD PMF calculations. These are higher than estimates of the association free energy of GpA in micelles, ranging from 15 to 38 kJ/mol. It is therefore apparent that differences in calculation methods can have as much of an effect on the estimate of ΔG°_{pmf} from computational methods as differences in hydrophobic environment and helix sequence.

5.5.2 EphA2: Two Interaction States?

In the ideal EphA2 PMF calculations, each minimum in the profiles corresponds to an interaction via either the glycine zipper or heptad motifs. As discussed in section 5.4.3, these profiles should be regarded as a qualitative description of the free energy landscape of the dimer, and indeed the glycine-restrained and heptad-restrained PMF calculations provide

[1] Sengupta and Marrink actually quote a value of -38.2 kJ/mol for ΔG_{pmf} and -29.2 kJ/mol for ΔG°_{pmf} (using a standard state of 1 M of micelle environment) without going into the details of the correction factor for the standard state. It is difficult to see how converting from a protein concentration of ~0.01 M to one of 1 M can result in a lower association free energy, so I assume that the sign of the correction factor applied to ΔG_{pmf} was incorrect and the appropriate value of ΔG°_{pmf} from the Sengupta profile is -47.2 kJ/mol. This is further suggested by the fact that the PMF depth obtained by Sengupta and Marrink is -40 kJ/mol, comparable with the results for the EphA2 dimers, using similar system sizes.

Table 5.2: Association free energies from PMF calculations. 'Model' refers to the secondary-structure restraints applied to the helices. 'Ideal' means the helices were restrained as ideal helices; 'NMR' means that the helix restraints were based on the NMR structure from Bocharov et al. [141]. Errors represent statistical errors from estimates from 500 ns bins of the simulation; none are quoted for the NMR PMFs as the simulation time was insufficient to calculate sufficient independent estimates of ΔG_{pmf}.

Helix	Model	Restraints	ΔG_{pmf} (kJ/mol)	ΔG°_{pmf} (kJ/mol)
EphA2	Ideal	Glycine zipper	-50.2 ± 1.9	-61.9
EphA2	Ideal	Heptad	-44.3 ± 1.8	-56.0
EphA2	NMR	Glycine zipper	-29	-42
EphA2	NMR	Heptad	-37	-50
EphA2	NMR	Cα	-39	-52
EphA1	Ideal	Glycine zipper	-40.9 ± 2.2	-52.5

mutually inconsistent estimates of the energy differences between the two interaction states. However, these PMF results, combined with the unrestrained simulations which resulted in states where the helices interacted via either the glycine zipper or heptad motifs, suggest that the TM region of the EphA2 receptor can interact via either the glycine zipper or the heptad motifs. It also seems likely that - in the CG model used - the interaction state via the glycine zipper motif is at least as stable as that via the heptad motif, and possibly more so. However, the large sources of possible error in what is still a new methodology mean that we will need new approaches to confidently quantify the free energy differences between the two interaction states. These approaches could involve larger computational resources allowing for much longer simulations; a similar PMF calculation done using atomistic techniques; a multi-dimensional PMF where the rotation of the helices is restrained as well as their separation; or a very large number of unrestrained simulations, sufficient to establish a probability distribution of interaction states. A confident description of the free energy landscape of the dimer system will probably require a combination of these approaches.

It is also significant that all PMF calculations were performed in a DPPC bilayer. The lipid environment is known to have a significant effect on the function of membrane proteins [250], and the Eph receptors are thought to exist in raft-like lipid microdomains [223, 224, 225]. The two postulated conformations of the EphA2 TM domain contain different hydrophobic lengths, so it is plausible that each conformation is stabilised by different lipid environments. Full characterisation of the energy states of EphA2 would require similar PMF calculations being performed in lipid bilayers with a variety of hydrophobic widths; we can speculate that the interaction via the glycine zipper motif, which seems most stable

in a DPPC bilayer, would be less stable relative to the interaction via the heptad motif in a thinner DMPC bilayer.

5.5.3 EphA2 vs EphA1 stability

EphA1 seems slightly less stable than EphA2, with a standard association free energy of 52.5±2.2 kJ/mol compared with 61.9±1.9 or 56.0±1.8 kJ/mol (the glycine zipper-restrained and heptad-restrained PMFs respectively). The error estimates of these numbers refer to the statistical error on a variety of estimates of the association energy from consecutive bins of simulation time. However, the true sources of error in these numbers are likely to stem from quantities which are difficult to quantify. These include insufficient sampling, the choice of secondary structure restraints, and a CG model which was parameterised using amino acid insertion into a hydrophobic environment, rather than amino acids interacting within a hydrophobic environment. It therefore seems reasonable to conclude that the two dimers are similarly stable, within the current resolution of this technique. No evidence of the EphA1 helices interacting via the alanine motif was seen. NMR spectroscopy suggested that the interaction via the alanine motif was stabilised relative to the interaction via the glycine zipper motif at (close to) physiological pH, when Glu^{547} is charged. Glu^{547} is charged in the simulations. Simulations using different bilayer thicknesses may again be postulated to affect the relative stability of the two possible interaction states, although there is no evidence as yet that an interaction via the alanine motif is at all stable in CGMD simulations.

5.6 Conclusions

PMF calculations performed using the EphA2 and EphA1 TM helices have shown that a fully converged profile requires at least 7500 ns of simulation time per simulation. The PMF profile was shown to be dependent on the flexibility allowed by the secondary structure restraints on the helices. Choosing different motif separations as the reaction coordinates produced free energy differences between equivalent states which were mutually inconsistent, suggesting sources of error beyond convergence issues. Despite these caveats, the simulations suggest that the EphA2 dimer can interact via a glycine zipper or a heptad motif, with the glycine zipper interaction at least as stable as the heptad interaction. The EphA1 dimer seems to be comparably stable to the EphA2 dimer, with no evidence of any interactions via the alanine motif. In conclusion, CG PMF calculations seem to be able to provide a useful qualitative picture of the energy landscape of a helix dimer system, but require careful

interpretation.

Chapter 6

Conclusions

6.1 Summary

This thesis has used MD simulations to investigate TM helices in bilayers. There have been two themes, complementing the two stage model of membrane protein folding: helix insertion into bilayers, and helix association within bilayers.

Chapter 3 discusses TM helix insertion into bilayers. This process was investigated using PMF calculations derived from CGMD simulations. The sequences investigated were chosen because data from translocon-mediated insertion experiments provided an understanding of the energetics of *in vivo* insertion for these sequences. The PMF profiles generated showed a significant dependence of TM helix insertion on both helix length and bilayer width. Consistent with several previous studies, the ΔG_{pmf} values were significantly larger than comparable values from the translocon-mediated insertion experiments. The sum of the per residue hydrophobicities of a sequence were compared with ΔG_{pmf} values, using a variety of hydrophobicity scales. The resulting ΔG^{add} values from this simple test showed that the ΔG_{pmf} values were consistent with hydrophobicity scales derived from all methods, except that from translocon-mediated insertion experiments. This suggests that the translocon-mediated insertion experiments measure something more complex than the free energy cost of taking a residue from an aqueous environment and placing it in a membrane. Instead, the interpretation of the experimental measurements depends on the insertion mechanism used by the translocon. The dependence of ΔG_{pmf} values on the width of the hydrophobic region of the bilayer provides support for one possible interpretation: if the effective free energy cost of moving a polypeptide chain into the translocon is zero - perhaps because it is tightly coupled with generating the chain within the ribosome - and the translo-

con selects which helices enter the membrane by some near-equilibrium process, then the experiments would measure the free energy cost of taking a residue from the translocon pore into the membrane. The experimental measurements would then become dependent on the environment of the translocon channel, including the hydrophobic pore ring.

Chapter 4 introduced the TM helices from the RTKs EphA1 and EphA2. The NMR structures for these helix dimers suggested that they could interact between two different motifs, possibly representing active and inactive states. Multiple CGMD simulations where the helices were allowed to dimerise in a lipid bilayer suggested that the EphA2 helices could interact via both glycine zipper and heptad motifs. However, the latter interaction produced a right-handed helix, in contradiction to the NMR structure. Simulations of the EphA1 helices showed a structure similar to the NMR structure, but provided no evidence of the helices interacting via the alanine motif. Atomistic simulations of the EphA2 helices suggested that neither structure was stable when represented in atomistic detail, although atomistic simulations of the EphA1 helices suggested that they did form a stable dimer.

Chapter 5 investigated the Eph receptor dimers further by performing PMF calculations on the dissociation of the helices. The calculated association free energy was larger than experimental measurements performed on comparable systems, although the dependence of the association on the environment makes direct comparisons difficult. The calculated association free energy was comparable with a previous CGMD PMF calculation on GpA. Similarly with the unrestrained simulations in Chapter 4, stable states were observed in the PMFs corresponding to EphA2 interacting via the glycine zipper and heptad motifs, and EphA1 interacting via the glycine zipper motif. No evidence of EphA1 interacting via the alanine motif was observed. The PMFs showed that convergence required at least 7.5 μs of simulation time per window, although the free energy difference between the two interaction modes of EphA2 was different when the PMF was calculated with different reaction coordinates, suggesting possible sources of error beyond convergence issues. The PMFs therefore should be interpreted as a semi-quantitative estimate of the free energy landscape of the system.

Overall, the simulations in this thesis have provided insights both into the function of some biological systems and the limitations of the CGMD model used. Results of PMF calculations of helix insertion into bilayers were consistent with a variety of other methods, suggesting the CG model used can reproduce free energies of insertion into bilayers at least semi-quantitatively. Comparison of these with experimental measurements taken *in vivo*

underlines the complexity of the cell membrane environment, and the care which must be taken when interpreting experimental results. CGMD simulations of the Eph receptor TM helices showed that care must also be taken when interpreting the results of simulations. Eph receptor dimers from CGMD simulations were not always consistent with NMR structures derived in a slightly different environment. This may be a consequence of the CG model used; however, given the previous success of the model in replicating helix-dimer contacts [135, 136] it is equally likely to be a consequence of different environments: lipid charge, hydrophobic width and interactions from other proteins all provide possible ways of stabilising a TM dimer.

6.2 Future Work

The work presented in this thesis could be usefully extended in a number of ways. These include: establishing a rigorous method of mapping free-energy landscapes of helix-bilayer systems; investigating the effect of the bilayer environment on free energy landscapes; extending existing methodologies to new systems; and investigating translocon function using direct simulations of the translocon.

To directly extend the work in this thesis, a priority would be to establish the reasons for two different implementations of WHAM providing different PMF profiles. An initial approach to doing so might be to test a variety of simple systems with both implementations, and progressively increase the complexity of the test systems until reasons for the difference can be established. Contacting the developers of both implementations, and perhaps trying to use other implementations, might also be useful.

The CG simulations of the TM regions of EphA1 and EphA2 have suggested a different possible interaction modes for both systems; however, the stability of each mode has not been conclusively established at a multiscale level. A range of simulations of both dimers at an atomistic level could establish with greater confidence the stability of the dimers. NMR spectroscopy and X-ray crystallography may perhaps be able to provide further structures of both systems, in multiple different conformations, which could provide a basis with which to compare dimers obtained from MD simulations. Similarly experimental data measuring the association free energy of the Eph receptor TM helices would provide a basis on which to assess the ability of the Bond CG model to provide quantitative measures of association free energies. It is important to compare MD simulation with comparable experimental data where possible, so experiments performed in pure DPPC would be particularly useful. An

alternative would be to attempt to perform MD simulations in bilayers which are a closer match to physiological membranes, although that would require knowledge of the composition of such membranes.

The PMF calculations in this chapter have shown that restraining one degree of freedom is not always sufficient to rigorously map the free energy landscape of a system. In the helix-bilayer PMFs, both obvious and less-obvious kinetic trapping of helices was observed, dependent on the starting structure of the helix. To be entirely confident of the free energy landscape of a helix interacting with a bilayer, a two-dimensional PMF should be calculated, where both the helix distance from the bilayer centre of mass and the helix tilt with respect to the bilayer normal are reaction coordinates. Care would need to be taken with such a calculation, as restraining the tilt of the helix may have implications for helix flexibility - for instance, some restraints may prevent a helix from kinking. Similarly a two-dimensional PMF calculation of a TM helix dimer, restraining the helix rotation angles as well as their separation, would ensure that the helices sample all possible interactions.

The simulations in this thesis have also implied that the bilayer environment has a significant effect on helices. Understanding the effect of bilayer width, headgroup charge and protein concentration on systems such as Eph receptors could be interesting. For example, it is possible that the EphA2 helices interact more stably via the heptad motif than the glycine zipper motif when the lipid bilayer is thinner than DPPC. Both multiscale simulations and PMF calculations can be used to develop a comprehensive understanding of helix-helix interactions. CG simulations need to be compared with experimental data and/or atomistic simulations to ensure that they can reproduce the physiological packing of a dimer; once this has been established, the long simulation times accessible to CG systems allow large numbers of such simulations to be performed, providing a possibility of identifying the effect of changing just one variable - such as bilayer width - on a system. Unrestrained CGMD simulations can be performed in large numbers to establish preferred interaction methods in different lipid environments; these can be converted to atomistic structures to establish the details of the interactions. PMF calculations, ideally in two or more dimensions, can provide at least a semi-quantitative view of the energy landscape. There are a large number of systems which could usefully use such methods, including other receptor tyrosine kinases and larger proteins such as bacterial outer membrane proteins. Including intra-and extra-cellular domains of RTKs in simulations could also provide further insights into the mechanisms of dimerisation.

The motivation for the helix-bilayer PMF calculations in this thesis was to investigate the insertion properties of the translocon. These calculations allowed us to make a prediction: that the free energy difference between taking a helix from water into a bilayer will be greater than the free energy difference of taking the same helix from the translocon interior into a bilayer. This prediction has already been validated to some extent by free energy perturbation calculations of a model poly-leucine helix [207]. Extending this approach of calculating the free energy change of taking a helix from the translocon into a bilayer to helices derived from proteins, rather than model helices, could provide more insights. Performing simulations of the translocon while including further elements of the experimental setup, such as the leader peptidase protein, could allow greater understanding of experimental results. Similarly simulations of polypeptide chains entering the translocon pore, perhaps with different forces applied, could provide more insights into the mechanism of translocon opening.

6.3 Conclusions

Computational methods have provided important insights into many aspects of biochemistry by providing a complementary approach to experiments and traditional theoretical techniques. This thesis has showed how quantitative free energy calculations can be applied to simulations of membrane systems, and has tested the extent to which coarse-grained models can produce quantitative information. These types of models, combined with increasing computer power and multiscale techniques, have the potential to yield much more information about membrane systems in the future.

Bibliography

[1] Bond PJ, Wee CL, Sansom MSP (2008) Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry* 47:11321–11331.

[2] Guidotti G (1972) Membrane proteins. *Annual Review of Biochemistry* 41:731–52.

[3] Hollán S (1996) Membrane fluidity of blood cells. *Haematologia (Budap)* 27:109–27.

[4] Lee AG (2003) Lipid-protein interactions in biological membranes: a structural perspective. *Biochim Biophys Acta* 1612:1–40.

[5] Marsh D (2008) Protein modulation of lipids, and vice-versa, in membranes. *Biochimica Et Biophysica Acta* 1778:1545–75.

[6] Simons K, Gerl MJ (2010) Revitalizing membrane rafts: new tools and insights. *Nat Rev Mol Cell Biol* 11:688–99.

[7] Simons K, Sampaio JL (2011) Membrane organization and lipid rafts. *Cold Spring Harb Perspect Biol* 3:a004697.

[8] Fagerberg L, Jonasson K, von Heijne G, Uhlén M, Berglund L (2010) Prediction of the human membrane proteome. *Proteomics* 10:1141–9.

[9] Dubois RN, et al. (1998) Cyclooxygenase in biology and disease. *Faseb Journal* 12:1063–1073.

[10] Robinson D, Wu Y, Lin S (2000) The protein tyrosine kinase family of the human genome. *Oncogene* 19:5548–5557.

[11] Kolakowski, Jr LF (1994) GCRDb: a G-protein-coupled receptor database. *Receptors Channels* 2:1–7.

[12] Verkman AS (2011) Aquaporins at a glance. *Journal of Cell Science* 124:2107–12.

[13] Jensen MØ, et al. (2010) Principles of conduction and hydrophobic gating in K+ channels. *Proceedings of the National Academy of Sciences USA* 107:5833–8.

[14] Holland IB, Blight MA (1999) ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *Journal of Molecular Biology* 293:381–99.

[15] Radestock S, Forrest LR (2011) The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of Molecular Biology* 407:698–715.

[16] Bordag N, Keller S (2010) α-helical transmembrane peptides: A "divide and conquer" approach to membrane proteins. *Chemistry and Physics of Lipids* 163:1–26.

[17] Du Plessis DJF, Nouwen N, Driessen AJM (2011) The Sec translocase. *Biochimica Et Biophysica Acta-Biomembranes* 1808:851–865.

[18] Rath A, Tulumello DV, Deber CM (2009) Peptide models of membrane protein folding. *Biochemistry* 48:3036–3045.

[19] von Heijne G (2011) Introduction to theme "membrane protein folding and insertion". *Annual Review of Biochemistry* 80:157–60.

[20] Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* 29:4031–7.

[21] Engelman DM, et al. (2003) Membrane protein folding: beyond the two stage model. *FEBS Letters* 555:122–5.

[22] van den Berg B, et al. (2004) X-ray structure of a protein-conducting channel. *Nature* 427:36–44.

[23] Tsukazaki T, et al. (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* 455:988–U72.

[24] Zimmer J, Nam Y, Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455:936–43.

[25] Egea PF, Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proceedings of the National Academy of Sciences USA* 107:17182–7.

[26] Driessen AJM, Nouwen N (2008) Protein translocation across the bacterial cytoplasmic membrane. *Annual Review of Biochemistry* 77:643–67.

[27] Ulbrandt N, Newitt J, Bernstein H (1997) The E-coli signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* 88:187–196.

[28] Connolly T, Gilmore R (1986) Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein. *Journal of Cell Biology* 103:2253–61.

[29] Tomkiewicz D, Nouwen N, Driessen AJM (2007) Pushing, pulling and trapping - modes of motor protein supported protein translocation. *Febs Letters* 581:2820–2828.

[30] Simon SM, Peskin CS, Oster GF (1992) What Drives the Translocation of Proteins? *Proceedings of the National Academy of Sciences* 89:3770–3774.

[31] Osborne AR, Rapoport TA, van den Berg B (2005) Protein translocation by the Sec61/SecY channel. *Annual Review of Cell Development Biology* 21:529–50.

[32] Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450:663–9.

[33] White S, von Heijne G (2004) The machinery of membrane protein assembly. *Current Opinion In Structural Biology* 14:397–404.

[34] White SH, von Heijne G (2008) How translocons select transmembrane helices. *Annual Review of Biophysics* 37:23–42.

[35] Hessa T, et al. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381.

[36] Hessa T, et al. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030.

[37] Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* 35:5109–5124.

[38] Goder V, Spiess M (2003) Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J* 22:3645–53.

[39] Young R, Bremer H (1976) Polypeptide-chain-elongation rate in escherichia coli b/r as a function of growth rate. *Biochem J* 160:185–94.

[40] Ilan J, Pierce DR, Hochberg AA, Folman R, Gyves MT (1984) Increased rates of polypeptide chain elongation in placental explants from human diabetics. *Proc Natl Acad Sci U S A* 81:1366–70.

[41] Varenne S, Knibiehler M, Cavard D, Morlon J, Lazdunski C (1982) Variable rate of polypeptide chain elongation for colicins a, e2 and e3. *J Mol Biol* 159:57–70.

[42] Zhang B, Miller, 3rd TF (2010) Hydrophobically stabilized open state for the lateral gate of the Sec translocon. *Proceedings of the National Academy of Sciences USA* 107:5399–404.

[43] MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: structure and implications. *Science* 276:131–3.

[44] Melnyk RA, Partridge AW, Deber CM (2002) Transmembrane domain mediated self-assembly of major coat protein subunits from ff bacteriophage. *Journal of Molecular Biology* 315:63–72.

[45] Dews IC, Mackenzie KR (2007) Transmembrane domains of the syndecan family of growth factor coreceptors display a hierarchy of homotypic and heterotypic interactions. *Proceedings of the National Academy of Sciences USA* 104:20782–7.

[46] Arkin IT, et al. (1994) Structural organization of the pentameric transmembrane α-helices of phospholamban, a cardiac ion channel. *EMBO J* 13:4757–64.

[47] Go MY, et al. (2006) Self-association of the transmembrane domain of an anthrax toxin receptor. *Journal of Molecular Biology* 360:145–56.

[48] Choma C, Gratkowski H, Lear JD, DeGrado WF (2000) Asparagine-mediated self-association of a model transmembrane helix. *Nature Structural Biology* 7:161–6.

[49] Dawson JP, Melnyk RA, Deber CM, Engelman DM (2003) Sequence context strongly modulates association of polar residues in transmembrane helices. *Journal of Molecular Biology* 331:255–62.

[50] Zhou FX, Merianos HJ, Brunger AT, Engelman DM (2001) Polar residues drive association of polyleucine transmembrane helices. *Proceedings of the National Academy of Sciences USA* 98:2250–5.

[51] Bocharov EV, et al. (2007) Unique dimeric structure of bnip3 transmembrane domain suggests membrane permeabilization as a cell death trigger. *Journal of Biological Chemistry* 282:16256–16266.

[52] Call ME, et al. (2006) The structure of the zeta zeta transmembrane dimer reveals features essential for its assembly with the t cell receptor. *Cell* 127:355–368.

[53] Choi MY, et al. (2005) Destabilization of the transmembrane domain induces misfolding in a phenotypic mutant of cystic fibrosis transmembrane conductance regulator. *Journal of Biological Chemistry* 280:4968–4974.

[54] Sal-Man N, Gerber D, Shai Y (2005) The identification of a minimal dimerization motif qxxs that enables homo- and hetero-association of transmembrane helices in vivo. *Journal of Biological Chemistry* 280:27449–27457.

[55] Bocharov EV, et al. (2008) Spatial structure of the dimeric transmembrane domain of the growth factor receptor erbb2 presumably corresponding to the receptor active state. *Journal of Biological Chemistry* 283:6950–6956.

[56] Stouffer AL, et al. (2008) Structural basis for the function and inhibition of an influenza virus proton channel (vol 451, pg 596, 2008). *Nature* 452:380–380.

[57] Mulholland AJ (2007) Chemical accuracy in qm/mm calculations on enzyme-catalysed reactions. *Chemistry Central Journal* 1:19.

[58] Senn HM, Thiel W (2009) Qm/mm methods for biomolecular systems. *Angew Chem Int Ed Engl* 48:1198–229.

[59] Strickland N, Mulholland AJ, Harvey JN (2006) The fe-co bond energy in myoglobin: a qm/mm study of the effect of tertiary structure. *Biophys J* 90:L27–9.

[60] van der Kamp MW, Perruccio F, Mulholland AJ (2008) High-level qm/mm modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. *Chem Commun (Camb)* pp 1874–6.

[61] Ridder L, Mulholland A, Rietjens I, Vervoort J (2000) A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase. *Journal of the American Chemical Society* 122:8728–8738.

[62] Roux B (2005) Ion conduction and selectivity in k+ channels. *Annual Review of Biophysics and Biomolecular Structure* 34:153–171.

[63] Roux B, et al. (2011) Ion selectivity in channels and transporters. *Journal of General Physiology* 137:415–426.

[64] Hub JS, De Groot BL (2008) Mechanism of selectivity in aquaporins and aquaglyceroporins. *Proceedings of the National Academy of Sciences of the United States of America* 105:1198–1203.

[65] Holyoake J, Sansom MSP (2007) Conformational change in an MFS protein: MD simulations of LacY. *Structure* 15:873–884.

[66] Pendse PY, Brooks BR, Klauda JB (2010) Probing the periplasmic-open state of lactose permease in response to sugar binding and proton translocation. *Journal of Molecular Biology* 404:506–521.

[67] Falconi M, et al. (2006) Structural dynamics of the mitochondrial ADP/ATP carrier revealed by molecular dynamics simulation studies. *Proteins-Structure Function and Bioinformatics* 65:681–691.

[68] Johnston JM, Khalid S, Sansom MSP (2008) Conformational dynamics of the mitochondrial ADP/ATP carrier: a simulation study. *Molecular Membrane Biology* 25:506–+.

[69] Celik L, Schiott B, Tajkhorshid E (2008) Substrate binding and formation of an occluded state in the leucine transporter. *Biophysical Journal* 94:1600–1612.

[70] Perilla JR, Beckstein O, Denning EJ, Woolf TB (2010) Computing ensembles of transitions from stable states: Dynamic importance sampling. *Journal of Computational Chemistry*.

[71] Leone V, Marinelli F, Carloni P, Parrinello M (2010) Targeting biomolecular flexibility with metadynamics. *Current Opinion In Structural Biology* 20:148–54.

[72] Delemotte L, Tarek M, Klein ML, Amaral C, Treptow W (2011) Intermediate states of the kv1.2 voltage sensor from atomistic molecular dynamics simulations. *Proceedings of the National Academy of Sciences USA* 108:6109–14.

[73] Schwaiger CS, Bjelkmar P, Hess B, Lindahl E (2011) 3-helix conformation facilitates the transition of a voltage sensor S4 segment toward the down state. *Biophysical Journal* 100:1446–54.

[74] Treptow W, Tarek M, Klein ML (2009) Initial response of the potassium channel voltage sensor to a transmembrane potential. *Journal of the American Chemical Society* 131:2107–9.

[75] Nury H, et al. (2010) One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue. *Proceedings of the National Academy of Sciences of the United States of America* 107:6275–6280.

[76] Dror RO, et al. (2009) Identification of two distinct inactive conformations of the β2-adrenergic receptor reconciles structural and biochemical observations. *Proceedings of the National Academy of Sciences of the United States of America* 106:4689–4694.

[77] Vanni S, Neri M, Tavernelli I, Rothlisberger U (2009) Observation of "ionic lock" formation in molecular dynamics simulations of wild-type beta(1) and beta(2) adrenergic receptors. *Biochemistry* 48:4789–4797.

[78] Zemel A, Ben-Shaul A, May S (2005) Perturbation of a lipid membrane by amphipathic peptides and its role in pore formation. *European Biophysics Journal* 34:230–42.

[79] May S (2002) Membrane perturbations induced by integral proteins: Role of conformational restrictions of the lipid chains. *Langmuir* 18:6356–6364.

[80] Wang J, Gambhir A, McLaughlin S, Murray D (2004) A computational model for the electrostatic sequestration of pi(4,5)p-2 by membrane-adsorbed basic peptides. *Biophysical Journal* 86:1969–1986.

[81] Atilgan AR, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal* 80:505–515.

[82] Keskin O, Jernigan RL, Bahar I (2000) Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophysical Journal* 78:2093–106.

[83] Lyman E, Pfaendtner J, Voth GA (2008) Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophysical Journal* 95:4183–4192.

[84] Tozzini V (2005) Coarse-grained models for proteins. *Current Opinion In Structural Biology* 15:144–150.

[85] Tirion (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 77:1905–1908.

[86] Zheng W, Brooks BR, Thirumalai D (2007) Allosteric transitions in the chaperonin groel are captured by a dominant normal mode that is most robust to sequence variations. *Biophys J* 93:2289–99.

[87] Zheng W, Liao JC, Brooks BR, Doniach S (2007) Toward the mechanism of dynamical couplings and translocation in hepatitis c virus ns3 helicase using elastic network model. *Proteins* 67:886–96.

[88] Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B* 108:750–760.

[89] Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP (2007) Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *Journal of Structural Biology* 157:593–605.

[90] Bond PJ, Sansom MSP (2006) Insertion and assembly of membrane proteins via simulation. *Journal of the American Chemical Society* 128:2697–2704.

[91] Shih AY, Arkhipov A, Freddolino PL, Schulten K (2006) Coarse grained protein-lipid model with application to lipoprotein particles. *Journal of Physical Chemistry B* 110:3674–84.

[92] Shih AY, Freddolino PL, Arkhipov A, Schulten K (2007) Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *Journal of Structural Biology* 157:579–92.

[93] Monticelli L, et al. (2008) The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation* 4:819–834.

[94] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B* 111:7812–7824.

[95] Baoukina S, Tieleman DP (2010) Direct simulation of protein-mediated vesicle fusion: lung surfactant protein b. *Biophysical Journal* 99:2134–42.

[96] Risselada HJ, Kutzner C, Grubmüller H (2011) Caught in the act: visualization of snare-mediated fusion events in molecular detail. *Chembiochem* 12:1049–55.

[97] Rzepiela AJ, Sengupta D, Goga N, Marrink SJ (2010) Membrane poration by antimicrobial peptides combining atomistic and coarse-grained descriptions. *Faraday Discussions* 144:431–43; discussion 445–81.

[98] Samuli Ollila OH, Louhivuori M, Marrink SJ, Vattulainen I (2011) Protein shape change has a major effect on the gating energy of a mechanosensitive channel. *Biophysical Journal* 100:1651–9.

[99] Sengupta D, Rampioni A, Marrink SJ (2009) Simulations of the c-subunit of ATP-synthase reveal helix rearrangements. *Molecular Membrane Biology* 26:422–34.

[100] DeVane R, Shinoda W, Moore PB, Klein ML (2009) A transferable coarse grain non-bonded interaction model for amino acids. *Journal of Chemical Theory and Computation* 5:2115–2124.

[101] Markvoort AJ, Pieterse K, Steijaert MN, Spijker P, Hilbers PAJ (2005) The bilayer-vesicle transition is entropy driven. *Journal of Physical Chemistry B* 109:22649–54.

[102] Markvoort AJ, Smeijers AF, Pieterse K, van Santen RA, Hilbers PAJ (2007) Lipid-based mechanisms for vesicle fission. *Journal of Physical Chemistry B* 111:5719–25.

[103] Smeijers AF, Pieterse K, Markvoort AJ, Hilbers PAJ (2006) Coarse-grained transmembrane proteins: hydrophobic matching, aggregation, and their effect on fusion. *Journal of Physical Chemistry B* 110:13614–23.

[104] Spijker P, et al. (2010) Coarse grained molecular dynamics simulations of transmembrane protein-lipid systems. *International Journal of Molecular Sciences* 11:2393–420.

[105] Orsi M, Haubertin DY, Sanderson WE, Essex JW (2008) A quantitative coarse-grain model for lipid bilayers. *Journal of Physical Chemistry B* 112:802–15.

[106] Orsi M, Michel J, Essex JW (2010) Coarse-grain modelling of DMPC and DOPC lipid bilayers. *Journal of Physics-Condensed Matter* 22:155106.

[107] Michel J, Orsi M, Essex JW (2008) Prediction of partition coefficients by multiscale hybrid atomic-level/coarse-grain simulations. *Journal of Physical Chemistry B* 112:657–60.

[108] Das A, Andersen HC (2009) The multiscale coarse-graining method. iii. a test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation. *Journal of Chemical Physics* 131:034102.

[109] Izvekov S, Voth GA (2005) A multiscale coarse-graining method for biomolecular systems. *Journal of Physical Chemistry B* 109:2469–73.

[110] Krishna V, Noid WG, Voth GA (2009) The multiscale coarse-graining method. iv. transferring coarse-grained potentials between temperatures. *Journal of Chemical Physics* 131:024103.

[111] Larini L, Lu L, Voth GA (2010) The multiscale coarse-graining method. vi. implementation of three-body coarse-grained potentials. *Journal of Chemical Physics* 132:164107.

[112] Noid WG, et al. (2008) The multiscale coarse-graining method. i. a rigorous bridge between atomistic and coarse-grained models. *Journal of Chemical Physics* 128:244114.

[113] Noid WG, et al. (2008) The multiscale coarse-graining method. ii. numerical implementation for coarse-grained molecular models. *Journal of Chemical Physics* 128:244115.

[114] Ayton GS, Voth GA (2010) Multiscale computer simulation of the immature hiv-1 virion. *Biophysical Journal* 99:2757–65.

[115] Hills, Jr RD, Lu L, Voth GA (2010) Multiscale coarse-graining of the protein energy landscape. *PLoS Comput Biol* 6:e1000827.

[116] Shi Q, Izvekov S, Voth GA (2006) Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane-bound ion channel. *Journal of Physical Chemistry B* 110:15045–8.

[117] Thorpe IF, Zhou J, Voth GA (2008) Peptide folding using multiscale coarse-grained models. *Journal of Physical Chemistry B* 112:13079–90.

[118] de Meyer FJM, Rodgers JM, Willems TF, Smit B (2010) Molecular simulation of the effect of cholesterol on lipid-mediated protein-protein interactions. *Biophysical Journal* 99:3629–38.

[119] de Meyer FJM, Venturoli M, Smit B (2008) Molecular simulations of lipid-mediated protein-protein interactions. *Biophysical Journal* 95:1851–1865.

[120] Venturoli M, Sperotto MM, Kranenburg M, Smit B (2006) Mesoscopic models of biological membranes. *Physics Reports-Review Section of Physics Letters* 437:1–54.

[121] Yiannourakou M, Marsella L, de Meyer F, Smit B (2010) Towards an understanding of membrane-mediated protein-protein interactions. *Faraday Discussions* 144:359–67; discussion 445–81.

[122] Som A, Vemparala S, Ivanov I, Tew GN (2008) Synthetic mimics of antimicrobial peptides. *Biopolymers* 90:83–93.

[123] Khandelia H, Ipsen JH, Mouritsen OG (2008) The impact of peptides on lipid membranes. *Biochimica Et Biophysica Acta-Biomembranes* 1778:1528–1536.

[124] Raghuraman H, Chattopadhyay A (2007) Melittin: A membrane-active peptide with diverse functions. *Bioscience Reports* 27:189–223.

[125] Li E, Hristova K (2006) Role of receptor tyrosine kinase transmembrane domains in cell signaling and human pathologies. *Biochemistry* 45:6241–6251.

[126] London E (2007) Using model membrane-inserted hydrophobic helices to study the equilibrium between transmembrane and nontransmembrane states. *Journal of General Physiology* 130:229–32.

[127] Killian JA (2003) Synthetic peptides as models for intrinsic membrane proteins. *FEBS Letters* 555:134–8.

[128] Im W, Chen J, Brooks, 3rd CL (2005) Peptide and protein folding and conformational equilibria: theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models. *Advances in Protein Chemistry* 72:173–98.

[129] Im W, Brooks, 3rd CL (2005) Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proceedings of the National Academy of Sciences USA* 102:6771–6.

[130] Ulmschneider MB, Doux JPF, Killian JA, Smith JC, Ulmschneider JP (2010) Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides. *Journal of the American Chemical Society* 132:3452–60.

[131] Ulmschneider MB, Smith JC, Ulmschneider JP (2010) Peptide partitioning properties from direct insertion studies. *Biophysical Journal* 98:L60–2.

[132] Ulmschneider JP, Andersson M, Ulmschneider MB (2011) Determining peptide partitioning properties via computer simulation. *Journal of Membrane Biology* 239:15–26.

[133] Kim T, Im W (2010) Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation. *Biophysical Journal* 99:175–83.

[134] Kandasamy S, Larson R (2006) Molecular dynamics simulations of model transmembrane peptides in lipid bilayers: A systematic investigation of hydrophobic mismatch. *Biophysical Journal* 90:2326–2343.

[135] Psachoulia E, Fowler PW, Bond PJ, Sansom MSP (2008) Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization. *Biochemistry* 47:10503–10512.

[136] Psachoulia E, Marshall DP, Sansom MSP (2010) Molecular dynamics simulations of the dimerization of transmembrane alpha-helices. *Accounts of Chemical Research* 43:388–396.

[137] Hall BA, Chetwynd AP, Sansom MSP (2011) Exploring peptide-membrane interactions with coarse-grained md simulations. *Biophysical Journal* 100:1940–8.

[138] Vostrikov VV, Hall BA, Greathouse DV, Koeppe, 2nd RE, Sansom MSP (2010) Changes in transmembrane helix alignment by arginine residues revealed by solid-state NMR experiments and coarse-grained MD simulations. *Journal of the American Chemical Society* 132:5803–11.

[139] Bond PJ, Parton DL, Clark JF, Sansom MSP (2008) Coarse-grained simulations of the membrane-active antimicrobial peptide maculatin 1.1. *Biophysical Journal* 95:3802–3815.

[140] Bocharov EV, et al. (2008) Spatial structure and ph-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase epha1. *Journal of Biological Chemistry* 283:29385–29395.

[141] Bocharov EV, et al. (2010) Left-handed dimer of epha2 transmembrane domain: Helix packing diversity among receptor tyrosine kinases. *Biophysical Journal* 98:881–889.

[142] Biggin PC, Bond PJ (2008) Molecular dynamics simulations of membrane proteins. *Methods in Molecular Biology* 443:147–160.

[143] Bond PJ, Sansom MSP (2007) Bilayer deformation by the kv channel voltage sensor domain revealed by self-assembly simulations. *Proceedings of the National Academy of Sciences USA* 104:2631–2636.

[144] Fowler PW, Tai K, Sansom MSP (2008) The selectivity of K^+ ion channels: testing the hypotheses. *Biophysical Journal* 95:5062–5072.

[145] Stansfeld PJ, Hopkinson R, Ashcroft FM, Sansom MSP (2009) Pip(2)-binding site in kir channels: definition by multiscale biomolecular simulations. *Biochemistry* 48:10926–10933.

[146] Born M, Oppenheimer R (1927) Zur quantentheorie der molekeln. *Annalen der Physik* 389:457–484.

[147] Hockney R (1970) The potential calculation and some applications.

[148] Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics* 76:637–649.

[149] Beeman D (1976) Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics* 20:130+.

[150] Morse PM (1929) Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Physical Review* 34:57–64.

[151] Leach A (2001) *Molecular Modelling: Principles and Applications (2nd Edition)* (Prentice Hall), 2 edition.

[152] van der Spoel D, et al. (2011) *Gromacs User Manual version 4.5.4*.

[153] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81:3684–3690.

[154] Nose S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* 52:255–268.

[155] Hoover WG (1985) Canonical dynamics - equilibrium phase-space distributions. *Physical Review A* 31:1695–1697.

[156] Nose S, Klein ML (1983) Constant pressure molecular-dynamics for molecular systems. *Molecular Physics* 50:1055–1076.

[157] Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals - a new molecular-dynamics method. *Journal of Applied Physics* 52:7182–7190.

[158] Darden T, York D, Pedersen L (1993) Particle mesh ewald: An n.log(n) method for ewald sums in large systems. *The Journal of Chemical Physics* 98:10089–10092.

[159] Essmann U, et al. (1995) A smooth particle mesh ewald method. *The Journal of Chemical Physics* 103:8577–8593.

[160] Kolling O (1996) Reaction field models for the solvent influence upon the fundamental carbonyl vibrational peak of 2-butanone in aprotic media. *Journal of Physical Chemistry* 100:16087–16091.

[161] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18:1463–1472.

[162] Sapay N, Tieleman DP (2010) Combination of the charmm27 force field with united-atom lipid force fields. *J Comput Chem*.

[163] DeVane R, et al. (2010) Parametrization and application of a coarse grained force field for benzene/fullerene interactions with lipids. *Journal of Physical Chemistry B* 114:16364–16372.

[164] Radzicka A, Wolfenden R (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* 27:1664–1670.

[165] Scott KA, et al. (2008) Coarse-grained md simulations of membrane protein-bilayer self-assembly. *Structure* 16:621–630.

[166] Torrie GM, Valleau JP (1974) Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters* 28:578 – 581.

[167] Roux B (1995) The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* 91:275 – 282.

[168] Aaqvist J (1990) Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* 94:8021–8024.

[169] Straatsma TP, Berendsen HJC (1988) Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of Chemical Physics* 89:5876–5886.

[170] Laio A, Parrinello M (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* 99:12562–12566.

[171] Darve E, Pohorille A (2001) Calculating free energies using average force. *Journal of Chemical Physics* 115:9169–9183.

[172] Park S, Araghi FK, Tajkhorshid E, Schulten K (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics* 119:3559–3566.

[173] Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. *Journal of Physical Chemistry B* 114:10235–10253.

[174] Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Physical Review Letters* 78:2690–2693.

[175] Chipot C, Henin J (2005) Exploring the free-energy landscape of a short peptide using an average force. *Journal of Chemical Physics* 123:244906.

[176] Henin J, Chipot C (2004) Overcoming free energy barriers using unconstrained molecular dynamics simulations. *Journal of Chemical Physics* 121:2904–2914.

[177] Van Der Spoel D, et al. (2005) Gromacs: fast, flexible, and free. *Journal of Computational Chemistry* 26:1701–1718.

[178] Marin M, Blondel M (2011) Editorial: In vivo protein folding - at the crossroad between basic research and biotechnology. *Biotechnology Journal* 6:615–617.

[179] von Heijne G (2007) Formation of transmembrane helices in vivo–is hydrophobicity all that matters? *Journal of General Physiology* 129:353–356.

[180] White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. *Annual Review of Biophysics and Biomolecular Structure* 28:319–365.

[181] Choe S, Hecht KA, Grabe M (2008) A continuum method for determining membrane protein insertion energies and the problem of charged residues. *Journal of General Physiology* 131:563–573.

[182] MacCallum JL, Bennett WFD, Tieleman DP (2007) Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *Journal of General Physiology* 129:371–377.

[183] Dorairaj S, Allen TW (2007) On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proceedings of the National Academy of Sciences USA* 104:4943–4948.

[184] Schow EV, et al. (2011) Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. *Journal of Membrane Biology* 239:35–48.

[185] Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157:105–132.

[186] Johansson ACV, Lindahl E (2009) The role of lipid composition for insertion and stabilization of amino acids in membranes. *Journal of Chemical Physics* 130.

[187] Johansson ACV, Lindahl E (2009) Titratable amino acid solvation in lipid membranes as a function of protonation state. *Journal of Physical Chemistry B* 113:245–53.

[188] Johansson ACV, Lindahl E (2009) Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proceedings of the National Academy of Sciences USA* 106:15684–9.

[189] Gkeka P, Sarkisov L (2010) Interactions of phospholipid bilayers with several classes of amphiphilic alpha-helical peptides: insights from coarse-grained molecular dynamics simulations. *Journal of Physical Chemistry B* 114:826–839.

[190] Gadsby DC, Vergani P, Csanady L (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* 440:477–483.

[191] Enquist K, et al. (2009) Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *Journal of Molecular Biology* 387:1153–1164.

[192] Alexander C, et al. (2009) Cystic fibrosis transmembrane conductance regulator: Using differential reactivity toward channel-permeant and channel-impermeant thiol-reactive probes to test a molecular model for the pore. *Biochemistry* 48:10078–10088.

[193] Doudou S, Burton NA, Henchman RH (2009) Standard free energy of binding from a one-dimensional potential of mean force. *Journal of Chemical Theory and Computation* 5:909–918.

[194] DeLano WL (2002) The pymol molecular graphics system.

[195] Grossfield A (2005) Wham: the weighted histogram analysis method.

[196] Jaud S, et al. (2009) Insertion of short transmembrane helices by the Sec61 translocon. *Proceedings of the National Academy of Sciences of the United States of America* 106:11588–11593.

[197] Vostrikov VV, Grant CV, Daily AE, Opella SJ, Koeppe, 2nd RE (2008) Comparison of "polarization inversion with spin exchange at magic angle" and "geometric analysis of labeled alanines" methods for transmembrane helix alignment. *Journal of the American Chemical Society* 130:12584–5.

[198] Junne T, Kocik L, Spiess M (2010) The Hydrophobic Core of the Sec61 Translocon Defines the Hydrophobicity Threshold for Membrane Integration. *Molecular Biology of the Cell* 21:1662–1670.

[199] Mitra K, Ubarretxena-Belandia T, Taguchi T, Warren G, Engelman D (2004) Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proceedings of the National Academy of Sciences of the United States of America* 101:4083–4088.

[200] Hedin LE, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *Journal of Molecular Biology* 396:221–9.

[201] MacCallum JL, Bennett WFD, Tieleman DP (2008) Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys J* 94:3393–3404.

[202] Wimley W, White S (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology* 3:842–848.

[203] Sharma V, et al. (2003) Crystal structure of Mycobacterium tuberculosis SecA, a preprotein translocating ATPase. *Proceedings of the National Academy of Sciences USA* 100:2243–8.

[204] Gumbart J, Chipot C, Schulten K (2011) Free energy of nascent-chain folding in the translocon. *Journal of the American Chemical Society*.

[205] Heinrich SU, Mothes W, Brunner J, Rapoport TA (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* 102:233–44.

[206] Jayasinghe S, Hristova K, White SH (2001) Energetics, stability, and prediction of transmembrane helices. *Journal of Molecular Biology* 312:927–34.

[207] Gumbart J, Chipot C, Schulten K (2011) Free-energy cost for translocon-assisted insertion of membrane proteins. *Proceedings of the National Academy of Sciences USA* 108:3596–3601.

[208] Cadena DL, Gill GN (1992) Receptor tyrosine kinases. *FASEB J* 6:2332–7.

[209] Finger C, Escher C, Schneider D (2009) The single transmembrane domains of human receptor tyrosine kinases encode self-interactions. *Science Signaling* 2:ra56.

[210] Adams RH, Klein R (2000) Eph receptors and ephrin ligands. essential mediators of vascular development. *Trends in Cardiovascular Medicine* 10:183–8.

[211] Himanen JP, Saha N, Nikolov DB (2007) Cell-cell signaling via eph receptors and ephrins. *Current Opinion in Cell Biology* 19:534–42.

[212] McBride JL, Ruiz JC (1998) Ephrin-a1 is expressed at sites of vascular development in the mouse. *Mechanisms of Development* 77:201–4.

[213] Pasquale EB (2005) Eph receptor signalling casts a wide net on cell behaviour. *Nature Reviews: Molecular Cell Biology* 6:462–75.

[214] Dodelet VC, Pasquale EB (2000) Eph receptors and ephrin ligands: embryogenesis to tumorigenesis. *Oncogene* 19:5614–9.

[215] Surawska H, Ma PC, Salgia R (2004) The role of ephrins and eph receptors in cancer. *Cytokine and Growth Factor Reviews* 15:419–33.

[216] Wykosky J, Debinski W (2008) The epha2 receptor and ephrina1 ligand in solid tumors: function and therapeutic targeting. *Molecular Cancer Research* 6:1795–806.

[217] Kullander K, Klein R (2002) Mechanisms and functions of eph and ephrin signalling. *Nature Reviews: Molecular Cell Biology* 3:475–86.

[218] Himanen JP, Nikolov DB (2003) Eph signaling: a structural view. *Trends in Neurosciences* 26:46–51.

[219] Himanen JP, et al. (2001) Crystal structure of an eph receptor-ephrin complex. *Nature* 414:933–8.

[220] Pitulescu ME, Adams RH (2010) Eph/ephrin molecules-a hub for signaling and endocytosis. *Genes & Development* 24:2480–2492.

[221] Egea J, Klein R (2007) Bidirectional eph-ephrin signaling during axon guidance. *Trends in Cell Biology* 17:230–8.

[222] Davy A, Aubin J, Soriano P (2004) Ephrin-b1 forward and reverse signaling are required during mouse development. *Genes and Development* 18:572–83.

[223] Brückner K, et al. (1999) Ephrinb ligands recruit grip family pdz adaptor proteins into raft membrane microdomains. *Neuron* 22:511–24.

[224] Gauthier LR, Robbins SM (2003) Ephrin signaling: One raft to rule them all? one raft to sort them? one raft to spread their call and in signaling bind them? *Life Sciences* 74:207–16.

[225] Wu C, Butz S, Ying Y, Anderson RG (1997) Tyrosine kinase receptors concentrated in caveolae-like domains from neuronal plasma membrane. *Journal of Biological Chemistry* 272:3554–9.

[226] Smith FM, et al. (2004) Dissecting the EphA3/Ephrin-A5 interactions using a novel functional mutagenesis screen. *Journal of Biological Chemistry* 279:9522–31.

[227] Wimmer-Kleikamp S, Janes P, Squire A, Bastiaens P, Lackmann M (2004) Recruitment of eph receptors into signaling clusters does not require ephrin contact. *Journal of Cell Biology* 164:661–666.

[228] Fleishman SJ, Schlessinger J, Ben-Tal N (2002) A putative molecular-activation switch in the transmembrane domain of erbb2. *Proceedings of the National Academy of Sciences of the United States of America* 99:15937–15940.

[229] Moriki T, Maruyama H, Maruyama I (2001) Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. *Journal of Molecular Biology* 311:1011–1026.

[230] Stansfeld PJ, Sansom MSP (2011) From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. *Journal of Chemical Theory and Computation* 7:1157–1166.

[231] Rotkiewicz P, Skolnick J (2008) Fast procedure for reconstruction of full-atom protein models from reduced representations. *Journal of Computational Chemistry* 29:1460–1465.

[232] Sali A, Blundell T (1993) Comparative protein modeling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234:779–815.

[233] Scott W, et al. (1999) The gromos biomolecular simulation program package. *Journal of Physical Chemistry A* 103:3596–3607.

[234] Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM (1984) A consistent empirical potential for water-protein interactions. *Biopolymers* 23:1513–1518.

[235] Kelley L, Gardner S, Sutcliffe M (1996) An automated approach for clustering an ensemble of nmr-derived protein structures into conformationally related subfamilies. *Protein Engineering* 9:1063–1065.

[236] Lee A (2004) How lipids affect the activities of integral membrane proteins. *Biochimica Et Biophysica Acta-Biomembranes* 1666:62–87.

[237] Fleming KG, Engelman DM (2001) Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A* 98:14340–4.

[238] Lee J, Im W (2008) Role of hydrogen bonding and helix-lipid interactions in transmembrane helix association. *Journal of the American Chemical Society* 130:6456–6462.

[239] Henin J, Pohorille A, Chipot C (2005) Insights into the recognition and association of transmembrane alpha-helices. the free energy of alpha-helix dimerization in glycophorin a. *Journal of the American Chemical Society* 127:8478–8484.

[240] Sengupta D, Marrink SJ (2010) Lipid-mediated interactions tune the association of glycophorin a helix and its disruptive mutants in membranes. *Physical Chemistry Chemical Physics* 12:12987–12996.

[241] Janosi L, Prakash A, Doxastakis M (2010) Lipid-modulated sequence-specific association of glycophorin a in membranes. *Biophysical Journal* 99:284–292.

[242] Prakash A, Janosi L, Doxastakis M (2010) Self-association of models of transmembrane domains of erbb receptors in a lipid bilayer. *Biophysical Journal* 99:3657–3665.

[243] Chng CP, Tan SM (2011) Leukocyte integrin l2 transmembrane association dynamics revealed by coarse-grained molecular dynamics simulations. *Proteins* 79:2203–13.

[244] Russ WP, Engelman DM (1999) TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proceedings of the National Academy of Sciences USA* 96:863–8.

[245] Doura AK, Fleming KG (2004) Complex interactions at the helix-helix interface stabilize the glycophorin a transmembrane dimer. *Journal of Molecular Biology* 343:1487–97.

[246] Doura AK, Kobus FJ, Dubrovsky L, Hibbard E, Fleming KG (2004) Sequence context modulates the stability of a gxxxg-mediated transmembrane helix-helix dimer. *Journal of Molecular Biology* 341:991–8.

[247] Cristian L, Lear JD, DeGrado WF (2003) Determination of membrane protein stability via thermodynamic coupling of folding to thiol-disulfide interchange. *Protein Science* 12:1732–40.

[248] Howard KP, Lear JD, DeGrado WF (2002) Sequence determinants of the energetics of folding of a transmembrane four-helix-bundle protein. *Proceedings of the National Academy of Sciences USA* 99:8568–72.

[249] Ebie AZ, Fleming KG (2007) Dimerization of the erythropoietin receptor transmembrane domain in micelles. *Journal of Molecular Biology* 366:517–24.

[250] Mackenzie KR (2006) Folding and stability of alpha-helical integral membrane proteins. *Chemical Reviews* 106:1931–77.

[251] MacKenzie KR, Fleming KG (2008) Association energetics of membrane spanning alpha-helices. *Current Opinion In Structural Biology* 18:412–419.

[252] Fisher LE, Engelman DM, Sturgis JN (2003) Effect of detergents on the association of the glycophorin A transmembrane helix. *Biophysical Journal* 85:3097–105.

[253] Zhang J, Lazaridis T (2006) Calculating the free energy of association of transmembrane helices. *Biophysical Journal* 91:1710–1723.

[254] Fleming KG, et al. (2004) Thermodynamics of glycophorin A transmembrane helix dimerization in C14 betaine micelles. *Biophysical Chemistry* 108:43–9.

[255] Fleming KG, Ackerman AL, Engelman DM (1997) The effect of point mutations on the free energy of transmembrane alpha-helix dimerization. *Journal of Molecular Biology* 272:266–75.

[256] Fleming KG (2002) Standardizing the free energy change of transmembrane helix-helix interactions. *Journal of Molecular Biology* 323:563–71.

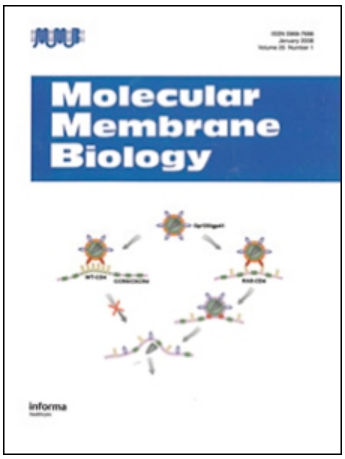
[257] Pongprayoon P, Beckstein O, Wee CL, Sansom MSP (2009) Simulations of anion transport through oprp reveal the molecular basis for high affinity and selectivity for phosphate. *Proceedings of the National Academy of Sciences USA* 106:21614–21618.

[258] Wee CL, Chetwynd A, Sansom MSP (2011) Membrane insertion of a voltage sensor helix. *Biophysical Journal* 100:410–419.

[259] Hub JS, de Groot BL, van der Spoel D (2010) A free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation* 6:3713–3720.

Appendices

This article was downloaded by: *[University of Oxford]*
On: *27 November 2008*
Access details: *Access Details: [subscription number 773573298]*
Publisher *Informa Healthcare*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Molecular Membrane Biology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713693962

CGDB: A database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations

Alan P. Chetwynd [a]; Kathryn A. Scott [b]; Younes Mokrab [a]; Mark S. P. Sansom [ab]
[a] Department of Biochemistry, University of Oxford, [b] Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, Oxford, UK

Online Publication Date: 01 December 2008

To cite this Article Chetwynd, Alan P., Scott, Kathryn A., Mokrab, Younes and Sansom, Mark S. P.(2008)'CGDB: A database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations',Molecular Membrane Biology,25:8,662 — 669
To link to this Article: DOI: 10.1080/09687680802446534
URL: http://dx.doi.org/10.1080/09687680802446534

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Molecular Membrane Biology, December 2008; 25(8): 662–669

informa
healthcare

CGDB: A database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations

ALAN P. CHETWYND[1], KATHRYN A. SCOTT[2], YOUNES MOKRAB[1], & MARK S. P. SANSOM[1,2]

[1]*Department of Biochemistry, University of Oxford, and* [2]*Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, Oxford, UK*

(Received 4 August 2008; and in revised form 25 August 2008)

Abstract
Membrane protein function and stability has been shown to be dependent on the lipid environment. Recently, we developed a high-throughput computational approach for the prediction of membrane protein/lipid interactions. In the current study, we enhanced this approach with the addition of a new measure of the distortion caused by membrane proteins on a lipid bilayer. This is illustrated by considering the effect of lipid tail length and headgroup charge on the distortion caused by the integral membrane proteins MscS and FLAP, and by the voltage sensing domain from the channel KvAP. Changing the chain length of lipids alters the extent but not the pattern of distortion caused by MscS and FLAP; lipid headgroups distort in order to interact with very similar but not identical regions in these proteins for all bilayer widths investigated. Introducing anionic lipids into a DPPC bilayer containing the KvAP voltage sensor does not affect the extent of bilayer distortion.

Keywords: *Coarse-grained, molecular dynamics, membrane protein*

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

Introduction

In the majority of genomes ~25% of open reading frames code for membrane proteins [1], which underlines their importance for cellular function. Membrane proteins function in the lipid environment of the membrane, and their insertion into the membrane must occur without compromising its effectiveness as a permeability barrier. It is therefore unsurprising that membrane proteins' structure and function is dependent on interactions with lipids. A number of different properties of the lipid bilayer have been shown experimentally to affect the activity of membrane proteins, including hydrophobic width, lipid phase and lipid headgroup structure [2–10]. A full understanding of membrane protein function therefore requires an understanding of the effect that the protein has on a lipid bilayer.

Molecular dynamics (MD) simulations can be used to model protein-lipid interactions, providing detailed information at the molecular level to complement experimental structural studies [11,12]. Atomistic approaches, which explicitly represent all atoms in a system, provide considerable detail [13,14] but with current computational resources are limited to time scales of ~100 ns, which limits the extent to which protein/lipid interactions may be sampled [15]. Meanwhile, coarse-grained (CG) methods provide access to longer time scales, and allow the simulation of larger systems [16–18]. A number of methods which treat the lipid bilayer as a hydrophobic slab have produced predictions of the position of membrane proteins in the bilayer [19–24]. However, these do not fully reproduce the complexity of the bilayer, particularly its deformability. Recently, we published a high-throughput self-assembly CGMD approach to predicting membrane protein position in the bilayer, allowing comparative analyses of protein/lipid interactions [25,26]. This paper presents an extension to this analysis focusing on how membrane lipid type affects the degree of bilayer deformation around the protein. This is relevant in the context of a number of experimental studies (reviewed in e.g., [8,27]; also see, e.g., [28]) of protein/lipid interactions and their functional relevance.

The effect of lipid tail length on the degree of distortion in the lipid bilayer is considered for two

Correspondence: Mark S. P. Sansom, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. E-mail: mark.sansom@bioch.ox.ac.uk

ISSN 0968-7688 print/ISSN 1464-5203 online © 2008 Informa UK Ltd
DOI: 10.1080/09687680802446534

example proteins: the 5-lipoxygenase activating protein (FLAP) is involved in the synthesis of leukotrienes, which binds to a lipid-soluble substrate (arachidonic acid) which is subsequently transferred to 5-lipoxygenase [29,30]; and the mechanosensitive channel of small conductance (MscS) [31], which is a well-characterized mechanosensitive channel which protects *Escherichia coli* from osmotic shock by opening in response to a sudden increase in transmembrane pressure [32]. FLAP was chosen owing to an unusual position of the binding site in the transmembrane region [33]; previous work (unpublished) on MscS had indicated that it caused notable distortion on a dipalmitoylphosphatidylcholine bilayer. In a third example, the voltage-sensor from the voltage-gated potassium channel KvaP was considered [34], by examining the effect of changing the lipid headgroup charge on the degree of bilayer distortion. The voltage sensor consists of four transmembrane segments and is unusual in that the fourth segment (S4) contains several basic amino acids, which act as primary elements of voltage sensing [35,36]. Both atomistic and CG simulations have suggested that the S4 contributes to local bilayer deformation which allows the charged residues to interact with polar headgroups and water [37–41].

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

Methods

The CG parameters for lipid and water molecules and Na^+ and Cl^- counterions have been described previously [16]. The CG parameters for amino acids were as described [42]. Briefly, one CG particle represents approximately four atoms and the protein backbone is represented by one particle with 0–3 side chain particles, depending on residue type. The CG representation of the protein was generated from the atomistic model obtained from X-ray diffraction or NMR crystallography. The proteins used were FLAP (pdb code 2Q7R), MscS (2OAU), and the voltage sensor (1ORS). The protein structure was maintained using an elastic network model with harmonic restraints applied between all backbone particles closer than 7 Å in the starting configuration. The lipids used to vary the bilayer width were dilauroylphosphatidylcholine (DLPC), dipalmitoylphosphatidylcholine (DPPC), and distearylphosphatidylcholine (DSPC), which have the same headgroup structure but have 12, 16 and 20 carbon atoms in each tail, respectively. The lipids used to vary the headgroup charge were DPPC and dipalmitoylphosphatidylglycerol (DPPG); these have identical tail lengths but DPPC has a zwitterionic headgroup whereas DPPG is anionic.

Following generation of the CG protein representation the system was prepared for simulation by the addition of randomly positioned CG lipid molecules followed by water and sufficient CG ions to ensure that the overall charge of the system was zero. Box sizes and the number of lipids used are listed in Table I. The system was subjected to fewer than 1000 steps of energy minimisation to relax any steric clashes.

All simulations were performed using GROMACS [43] (www.gromacs.org). Lennard-Jones interactions were smoothly shifted to zero between 9 and 12 Å, and electrostatic interactions were shifted to zero between 0 and 12 Å. Simulations were performed at constant pressure, temperature and number of particles; the pressure was 1 bar and the temperature 323 K. The temperature of the protein, lipid and solvent were coupled separately using the Berendsen algorithm [44]. A coupling constant of $\tau_p = 1$ ps was used, with compressibility of 5×10^{-6} bar^{-1}. Timesteps were 20–40 fs.

Analysis was performed only on the last 100 ns of each simulation to allow for equilibration. In the analysis of bilayer distortion the position of the protein was fixed in the x-y plane allowing identification of the areas of the protein which caused specific distortion. 5×5 $Å^2$ bins were defined relative to the centre of mass of the protein. The distance of the phosphate headgroups above or below the bilayer centre of mass was recorded for each bin every 400 ps, and the average for each bin was taken over the period analysed (described above). Analysis was performed using tools that are part of the GROMACS package and in-house software.

Results

The coarse-grain database

An overview of the data in the coarse-grain database (CGDB, http://sbcb.bioch.ox.ac.uk/cgdb/index.php) has been presented elsewhere [25,26]. The aim of

Table I. Lipid types, lipid numbers and box sizes used in simulations.

Protein	Lipid type	Lipid number	Box size ($Å^3$)
FLAP	DLPC	280	$120 \times 120 \times 130$
FLAP	DPPC	319	$120 \times 120 \times 160$
FLAP	DSPC	274	$120 \times 120 \times 130$
MscS	DLPC	418	$120 \times 120 \times 180$
MscS	DPPC	387	$150 \times 150 \times 150$
MscS	DSPC	331	$110 \times 110 \times 210$
Voltage sensor	DPPC	320	$100 \times 100 \times 180$
Voltage sensor	DPPC:DPPG	270:30	$100 \times 100 \times 130$
Voltage sensor	DPPC:DPPG	258:43	$100 \times 100 \times 150$
Voltage sensor	DPPC:DPPG	225:75	$100 \times 100 \times 170$
Voltage sensor	DPPC:DPPG	150:150	$100 \times 100 \times 150$

this paper is to describe an extension of the analysis focusing on the perturbation of the lipid bilayer seen with a number of different lipids. Briefly, we discuss the data set available in the CGDB, and illustrate how it may be used by considering the effect of changing the lipid tail length on two example proteins, FLAP and MscS. We also consider the effect of changing the lipid headgroup charge on the distortion caused by the voltage sensor.

The CGDB contains predictions of the position of membrane proteins in the lipid bilayer, currently for 126 protein structures. For each simulation the following data is provided: (i) An image of the final snapshot of each simulation showing the protein's backbone structure and the lipid headgroups; (ii) the coordinates of the final frame in the simulation; (iii) graphs showing the root mean square deviation of the backbone C-alpha particles from the starting structure as a function of time; graphs showing the root mean square fluctuation of each residue away from its average position over the simulation; (iv) a graphical representation of the protein with each residue coloured according to the amount of the simulation time it interacts with the lipid headgroups; and (v) a table containing data for each residue, including the fraction of the simulation time the residue interacts with lipid headgroups and tails, the total number of contacts each residue makes with lipid headgroups and tails and the height of the residue in the plane of the bilayer (Figure 1). In addition, contour plots are calculated to illustrate the distance of the phosphate headgroups above and below the bilayer centre of mass in the upper and

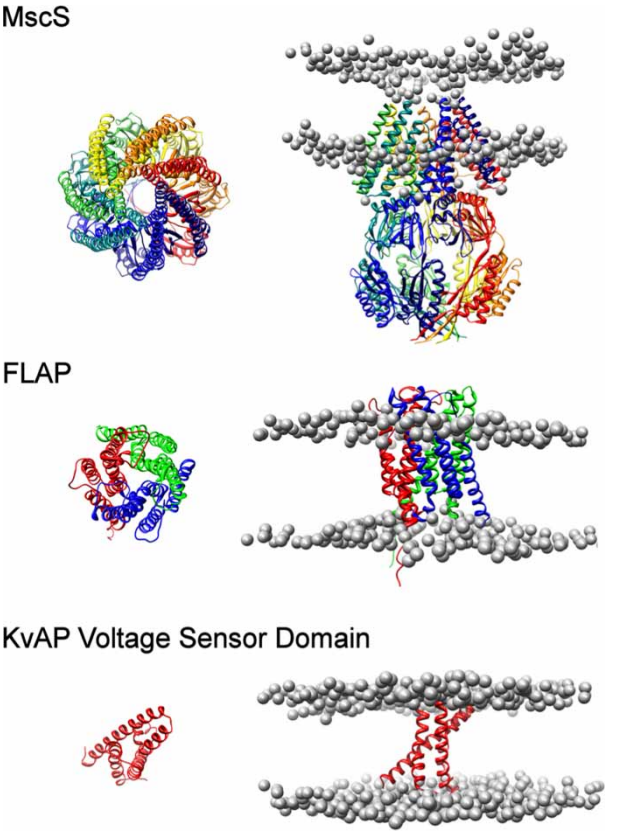


Figure 2. Positions and orientations of MscS, FLAP and the KvAP sensor domain. The position of MscS (top), FLAP (middle) and KvAP voltage sensor (bottom) in a DPPC bilayer. The protein structure is shown as a ribbon diagram generated from the original pdb structure following superposition with the simulated coarse-grained structure. Phosphate particles are shown as grey spheres.

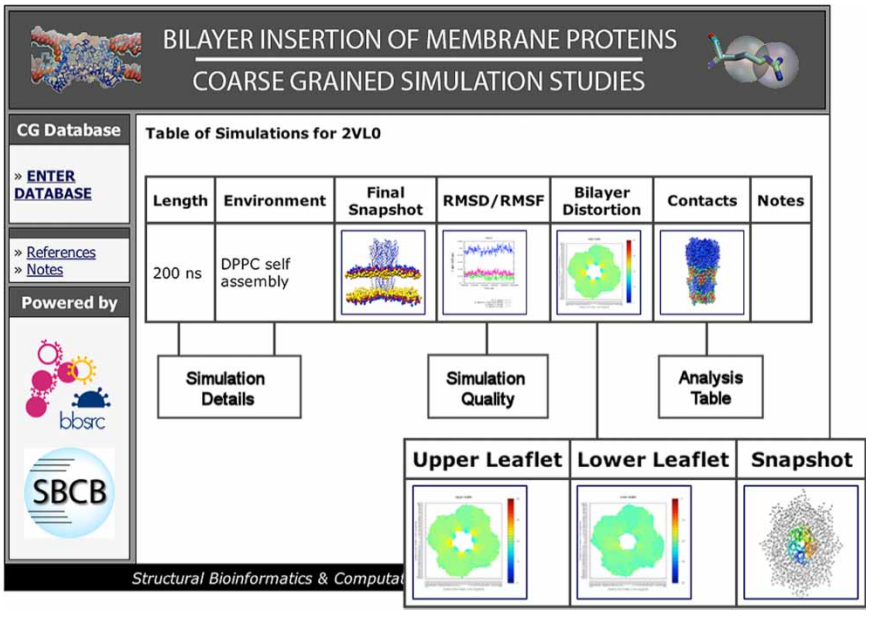


Figure 1. The Coarse-Grain Database. This Figure shows a screen shot of the database entry for a ligand gated ion channel from *E. chrysanthemi* [50] pdb code 2VL0. Each of the picture 'tabs' provides a link to downloadable data files.

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

lower leaflets, allowing identification of the regions of the protein that cause significant distortion of the lipid bilayer. The data are available for download, allowing further analysis and comparison of sets of proteins.

The effect of lipid tail length on bilayer perturbation

To investigate the effect of lipid tail length on distortion we ran simulations of MscS and FLAP (Figure 2) with three different lipids: DLPC, which has three particles in each tail in the CG model, DPPC, which has four, and DSPC, which has five tail particles.

For the simulations involving MscS, the patterns of distortion were the same for all lipid types (Figure 3). In the upper leaflet, the distortion is seen at radial distances up to ~70 Å away from the protein's centre of mass. The phosphate headgroups are approximately 5 Å away from the bilayer centre of mass for DLPC, DPPC and DSPC, with the headgroup height above the bilayer centre away from the protein ~22 Å, ~25 Å and ~28 Å for DLPC, DPPC and DSPC respectively. This is largely due to a ring of seven tyrosine residues (Tyr 27 from the inside tip of each monomer of the MscS heptamer) which interacts with DPPC headgroups for 97% of the simulation. It has been suggested that tyrosine interacts with the linking glycerol particles in lipid bilayers, acting as anchors in the interfacial region of membrane proteins [45]. An adjacent ring of arginine residues (Arg 88) are somewhat less exposed to the lipid headgroups (interact with DPPC headgroups for 48% of the analysed simulation). In the lower leaflet, the phosphate headgroups are drawn down to interact with the protein to approximately 30 Å away from the bilayer centre of mass. This is due to a ring of seven arginine residues (Arg 59); of these, six interact strongly with the DLPC headgroups, five with the DPPC headgroups and four with the DSPC headgroups (for 96%, 98% and 99% of the analysed simulation time, respectively).

Radial plots of phosphate distance away from the bilayer centre indicate that the headgroups interact with similar, though not identical, areas of the protein for all lipid tail lengths (Figure 4). For the longer lipid tails the degree of distortion is higher than for the shorter tail lengths, although the areas of the protein with which the DSPC headgroups interact are ~2–3 Å further away from the bilayer centre than the areas with which the DLPC headgroups interact. This suggests that the adjustment for hydrophobic mismatch is greater with longer lipid tail lengths, but is not complete.

For FLAP the areas of distortion extend to ~50 Å radially away from the protein's centre of mass (Figure 5). The protein is located relatively high in the upper leaflet, which is associated with a high density of charged residues, particularly arginines, in the upper third of the protein. Some of these arginine side chains are 'snorkelling': reaching up

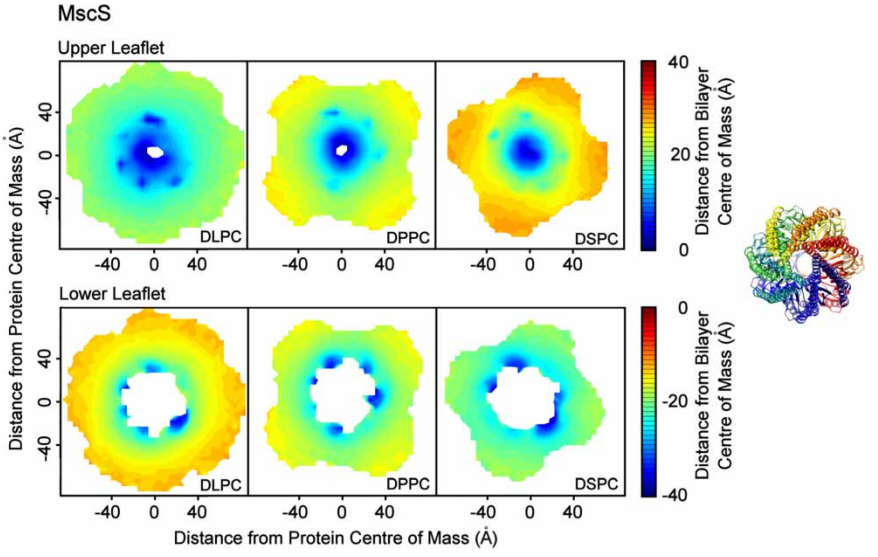


Figure 3. Bilayer perturbation around MscS. Contour plots (on the left hand side) illustrating the change in the extent of the perturbation of the lipid bilayer around MscS with lipid tail length. The right hand side shows the orientation of the protein. The data were calculated using a set of structures with the protein superposed in the x-y plane such that the protein had the same orientation in all cases. The data were split into 5 × 5 Å bins for analysis. Smoothed contour plots are coloured according to the mean distance of the phosphate headgroups from the bilayer centre of mass in each bin. The variation in the data was similar in each bin, with standard deviation values typically in the range of 2–3 Å. The data show that the pattern of bilayer distortion is similar for different lipid types with the extent depending upon the hydrophobic mismatch. The upper leaflets for each lipid show isotropic distortion. The lower leaflets for each lipid show distinct areas of distortion caused by Arg 59.

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

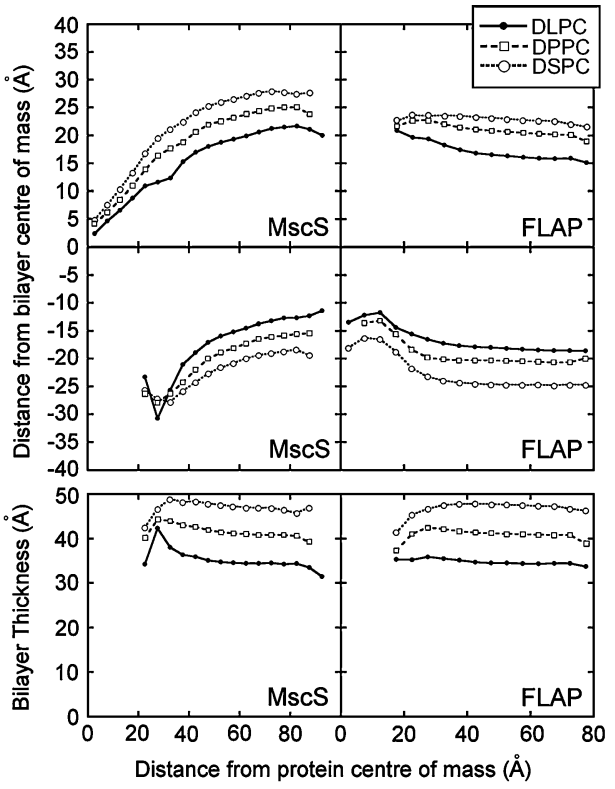


Figure 4. Bilayer distortion as a function of distance from the protein. The upper four panels show the mean phosphate radial distance from the bilayer centre of mass in the upper and lower leaflets as a function of the distance from the protein centre of mass of MscS and FLAP. In each case the pattern of bilayer distortion is similar irrespective of lipid tail length. The lower two panels give the bilayer thickness as function of the distance from the protein centre of mass.

to interact with the lipid headgroup/water interface region. The lower leaflet indicates two main areas of distortion, although FLAP is a trimer. One subunit in the structure is missing eight residues; it is this space into which the lipid headgroups are drawn. In the upper leaflet, two residues in each subunit (Lys 116 and Arg 117) interact heavily with the DPPC headgroups (for 96% and 94% of the analysed simulation, respectively), drawing them down while four residues (Phe 25, Lys 96, Gln 95, Phe 92) interact with the glycerol backbone of the lipids (for 85%, 84%, 83% and 79%, respectively), ensuring that the lipids interact slightly higher with the rest of the protein.

The radial plots of phosphate distance away from the bilayer centre indicate that in the upper leaflet of the DLPC bilayer the average phosphate height changes from ~16 Å around 60 Å away from the protein to ~21 Å at around 20 Å from the protein (Figure 4). In contrast, the phosphate headgroups of DSPC remain at ~23 Å above the bilayer centre. In the lower leaflet there is significant distortion in all simulations, with the headgroups ~18 Å below the bilayer centre at 60 Å away from the protein in the DLPC simulation but only ~12 Å below at 5 Å from the protein. In the DSPC simulation the phosphate headgroups are on average ~25 Å below the bilayer centre at 60 Å from the protein, but only ~16 Å below at 5 Å away from the protein. This is a larger adjustment than in the DLPC simulation, although the adjustment in the upper leaflet was larger for DLPC than DSPC. These results suggest that is the optimal bilayer thickness for FLAP is ~41 Å wide, similar to DPPC.

The effect of lipid headgroup charge on bilayer distortion

A number of simulation studies (see above) have indicated interactions between the voltage sensor of Kv channels and lipid headgroups. These are likely to be of functional importance [36] and have subsequently been seen in a crystal structure of a Kv channel [46]. To date, most simulation studies have focussed on zwitterionic lipids (i.e., phosphatidylycholines). To investigate the effect of lipid headgroup charge on bilayer distortion, five simulations were run of the voltage sensor in lipid bilayers, using DPPC:DPPG lipid molar ratios of 1:0, 9:1, 6:1, 3:1 and 1:1. The patterns of distortion for all simulations were the same for all simulations. The extent of the distortion was not affected by changing the charge of the lipid headgroup (Figure 6). In the upper leaflet, the phosphate particles move from ~20 Å above the bilayer centre at 60 Å away from the protein to ~16–18 Å above the bilayer centre at 8 Å away from the protein. While the DPPC-only bilayer shows the least distortion, the 9:1 DPPC:DPPG mixed bilayer shows the largest distortion, all variation being within typical standard deviations of the data set (~5 Å). In the lower leaflet, the average position of the phosphate headgroups shifts from ~21 Å below the bilayer centre of mass at 60 Å away from the protein to ~18 Å below at 8 Å away from the protein. Therefore, as a whole, each simulation shows a very similar level of distortion and there is no correlation between the extent of the distortion caused by the voltage sensor and the concentration of anionic lipids in the bilayer. This suggests that the voltage sensor interaction is relatively insensitive to the lipid headgroup, provided there is a negatively charged phosphate group present. This correlates well with reconstitution experiments [36].

Discussion and conclusion

Interactions with lipids have been shown to be important for the structure and function of a variety of membrane proteins. The insertion of membrane

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

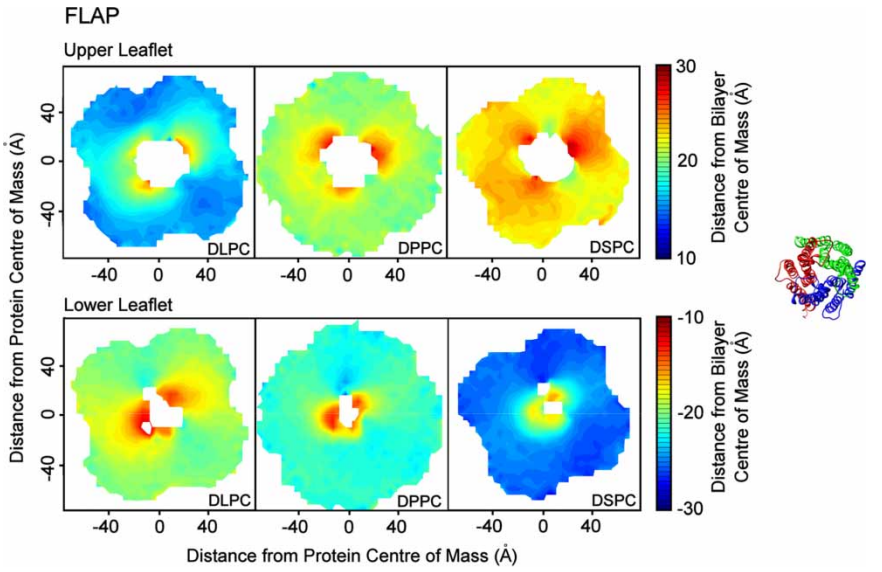


Figure 5. Bilayer perturbation around FLAP. Contour plots (on the left hand side) showing the change in the extent of the perturbation of the lipid bilayer around FLAP with lipid tail length. The right hand side shows the orientation of the protein. The analysis was performed as described for MscS. The patterns of distortion are similar for all lipid types, with the extent dependent on the bilayer width. The upper leaflets for each lipid show three areas where the headgroups are drawn up (red sections). The lower leaflets for each lipid show only two areas where the headgroups are drawn up (red sections) as one chain in the structure of FLAP is missing eight residues.

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

proteins into a bilayer can therefore cause significant distortion to the bilayer, and investigating this distortion can shed light on the function of the protein. The data in the CGDB are enhanced by consideration of this distortion. The CGMD simulation approach used to generate CGDB entries can be completed for a new membrane protein structure within hours/days, thus providing a model of protein/lipid interactions which can form a starting point for more detailed atomistic MD simulations and/or detailed biophysical investigations.

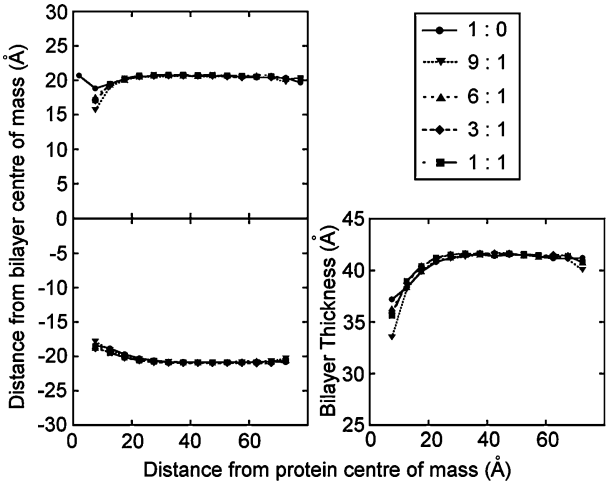


Figure 6. Bilayer distortion by the KvAP voltage sensor. The mean phosphate distance from the bilayer centre of mass in the (A) upper and (B) lower leaflets is shown as a function of distance from the protein surface for the voltage sensor (pdb code 1ORS). The bilayer (phosphate-phosphate) thickness is shown in (C). Five simulations were run, with a DPPC:DPPG ratio of 1:0, 9:1, 6:1, 3:1 and 1:1. The extent of the distortion is similar in all simulations, with no correlation between the extent of the distortion and the concentration of anionic lipids.

The investigation of the impact of different widths of bilayer on two example proteins showed that the extent of the distortion is dependent on lipid tail length, as the bilayer adapts to allow the headgroups to interact with the same areas of the protein. Comparable behaviour has also been seen in CG simulations of rhodopsin [47]. The lipid headgroups interact with very similar, yet non-identical regions of the protein despite having bilayer widths varying from ~35 Å to ~45 Å. This suggests that the energetic cost of distorting the bilayer is generally less than the energetic cost of unfavourable interactions between lipid headgroups and the protein surface. Investigation into the effect of introducing anionic lipids on the distortion caused by the voltage sensor indicates that the level of distortion is unaffected by the presence or concentration of DPPG in a DPPC bilayer, despite the presence of a number of basic residues in the centre of the bilayer.

It has been proposed that, in the case of significant hydrophobic mismatch, some proteins may distort to match the hydrophobic length of the bilayer, rather than the bilayer distorting to match the hydrophobic length of the protein (for a recent study, see e.g., [48]). Our results indicate that the bilayer can distort to match the hydrophobic length of the protein at a range of bilayer thicknesses. Nevertheless, an elastic network model [49] was used in the current study in

order to restrain the protein to ensure its stability during the simulation, which may prevent it from distorting unduly to adapt to the hydrophobic length of the lipid bilayer. An extension to this work would be to consider similar systems with unrestrained proteins in atomistic simulations.

Acknowledgements

This work was supported by the BBSRC, via the Membrane Protein Structure Initiative, and the Oxford Centre of Integrative Systems Biology. A.C. is a Medical Research Council student.

***Declaration of interest*:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

[1] Wallin E, von Heijne G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archean, and eukaryotic organisms. Prot Sci 7:1029–1038.

[2] Cornelius F. 2001. Modulation of Na,K-ATPase and Na-ATPase activity by phospholipids and cholesterol. I. Steady-state kinetics. Biochemistry 40:8842–8851.

[3] Starling AP, Khan YM, East JM, Lee AG. 1994. Characterization of the single Ca^{2+}-binding site on the Ca^{2+}-ATPase reconstituted with short-chain or long-chain phosphatidylcholines. Biochem J 304:569–575.

[4] Sunshine C, McNamee MG. 1994. Lipid modulation of nicotinic acetylcholine-receptor function – the role of membrane lipid-composition and fluidity. Biochim Biophys Acta 1191:59–64.

[5] Baldwin PA, Hubbell WL. 1985. Effects of lipid environment on the light-induced conformational-changes of rhodopsin.1. Absence of metarhodopsin-II production in dimyristoylphosphatidylcholine recombinant membranes. Biochemistry 24:2624–2632.

[6] Baldwin PA, Hubbell WL. 1985. Effects of lipid environment on the light-induced conformational-changes of rhodopsin. 2. Roles of lipid chain-length, unsaturation, and phase state. Biochemistry 24:2633–2639.

[7] Caffrey M, Feigenson GW. 1981. Fluorescence quenching in model membranes. 3. Relationship between calcium adenosine-triphosphatase enzyme-activity and the affinity of the protein for phosphatidylcholines with different acyl chain characteristics. Biochemistry 20:1949–1961.

[8] Lee AG. 2004. How lipids affect the activities of integral membrane proteins. Biochim Biophys Acta 1666:62–87.

[9] Lange C, Nett JH, Trumpower BL, Hunte C. 2001. Specific roles of protein-phospholipid interactions in the yeast cytochrome bc(1) complex structure. EMBO J 20:6591–6600.

[10] Fyfe PK, Isaacs NW, Cogdell RJ, Jones MR. 2004. Disruption of a specific molecular interaction with a bound lipid affects the thermal stability of the purple bacterial reaction centre. Biochim Biophys Acta 1608:11–22.

[11] Fyfe PK, McAuley KE, Roszak AW, Isaacs NW, Codgell RJ, Jones MR. 2001. Probing the interface between membrane proteins and membrane lipids by X-ray crystallography. Trends Biochem Sci 26:106–112.

[12] Hunte C, Richers S. 2008. Lipids and membrane protein structures. Curr Opin Struct Biol 18:406–411.

[13] Feller SE, Gawrisch K, Woolf TB. 2003. Rhodopsin exhibits a preference for solvation by polyunsaturated docosohexaenoic acid. J Am Chem Soc 125:4434–4435.

[14] Grossfield A, Feller SE, Pitman MC. 2006. A role for direct interactions in the modulation of rhodopsin by omega-3 polyunsaturated lipids. Proc Natl Acad Sci USA 103:4888–4893.

[15] Grossfield A, Feller SE, Pitman MC. 2007. Convergence of molecular dynamics simulations of membrane proteins. Proteins: Struct Func Bioinf 67:31–40.

[16] Marrink SJ, de Vries AH, Mark AE. 2004. Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108:750–760.

[17] Marrink SJ, Risselada J, Yefimov S, Tieleman DP, de Vries AH. 2007. The MARTINI forcefield: coarse grained model for biomolecular simulations. J Phys Chem B 111:7812–7824.

[18] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. 2008. The MARTINI coarse grained force field: extension to proteins. J Chem Theor Comp 4:819–834.

[19] Basyn F, Charloteaux B, Thomas A, Brasseur R. 2001. Prediction of membrane protein orientation in lipid bilayers: a theoretical approach. J Mol Graphics Modelling 20:235–244.

[20] Basyn F, Spies B, Bouffioux O, Thomas A, Brasseur R. 2003. Insertion of X-ray structures of proteins in membranes. J Mol Graphics Modelling 22:11–21.

[21] Ducarme P, Rahman N, Brasseur R. 1998. IMPALA: a simple restraint field to simulate the biological membrane in structure studies. Prot Struct Func Genet 30:357–371.

[22] Im W, Feig M, Brooks CL. 2003. An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. Biophys J 85:2900–2918.

[23] Ulmschneider MB, Sansom MSP, Di Nola A. 2005. Properties of integral membrane protein structures: derivation of an implicit membrane potential. Proteins: Struct Funct Bioinf 59:252–265.

[24] Ulmschneider MB, Sansom MSP, Di Nola A. 2006. Evaluating tilt angles of membrane-associated helices: comparison of computational and NMR techniques. Biophys J 90:1650–1660.

[25] Sansom MSP, Scott KA, Bond PJ. 2008. Coarse grained simulation: a high throughput computational approach to membrane proteins. Biochem Soc Transac 36:27–32.

[26] Scott KA, Bond PJ, Ivetac A, Chetwynd AP, Khalid S, Sansom MSP. 2008. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. Structure 16:621–630.

[27] Andersen OS, Koeppe RE. 2007. Bilayer thickness and membrane protein function: an energetic perspective. Ann Rev Biophys Biomol Struct 36:107–130.

[28] Powl AM, East JM, Lee AG. 2007. Different effects of lipid chain length on the two sides of a membrane and the lipid annulus of MscL. Biophys J 93:113–122.

[29] Dixon RAF, Diehl RE, Opas E, Rands E, Vickers PJ, Evans JF, Gillard JW, Miller DK. 1990. Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. Nature 343:282–284.

[30] Miller DK, Gillard JW, Vickers PJ, Sadowski S, Leveille C, Mancini JA, Charleson P, Dixon RAF, Fordhutchinson AW, Fortin R, Gauthier JY, Rodkey J, Rosen R, Rouzer C, Sigal IS, Strader CD, Evans JF. 1990. Identification and isolation of a membrane-protein necessary for leukotriene production. Nature 343:278–281.

Downloaded By: [University of Oxford] At: 15:53 27 November 2008

[31] Steinbacher S, Bass R, Strop P, Rees DC. 2007. Structures of the prokaryotic mechanosensitive channels MscL and MscS. In: Mechanosensitive Ion Channels, Part A pp 1–24.
[32] Edwards MD, Booth IR, Miller S. 2004. Gating the bacterial mechanosensitive channels: MscS a new paradigm? Curr Opin Microbiol 7:163–167.
[33] Ferguson AD, McKeever BM, Xu SH, Wisniewski D, Miller DK, Yamin TT, Spencer RH, Chu L, Ujjainwalla F, Cunningham BR, Evans JF, Becker JW. 2007. Crystal structure of inhibitor-bound human 5-lipoxygenase-activating protein. Science 317:510–512.
[34] Jiang Y, Lee A, Chen J, Ruta V, Cadene M, Chait BT, Mackinnon R. 2003. X-ray structure of a voltage-dependent K^+ channel. Nature 423:33–41.
[35] Ramsey IS, Moran MM, Chong JHA, Clapham DE. 2006. A voltage-gated proton-selective channel lacking the pore domain. Nature 440:1213–1216.
[36] Schmidt D, Jiang QX, MacKinnon R. 2006. Phospholipids and the origin of cationic gating charges in voltage sensors. Nature 444:775–779.
[37] Freites JA, Tobias DJ, von Heijne G, White SH. 2005. Interface connections of a transmembrane voltage sensor. Proc Natl Acad Sci USA 102:15059–15064.
[38] Treptow W, Tarek M. 2006. Environment of the gating charges in the Kv1.2 Shaker potassium channel. Biophys J 90:L64–66.
[39] Bond PJ, Sansom MSP. 2007. Bilayer deformation by the Kv channel voltage sensor domain revealed by self-assembly simulations. Proc Natl Acad Sci USA 104:2631–2636.
[40] Sands ZA, Sansom MSP. 2007. How does a voltage-sensor interact with a lipid bilayer? Simulations of a potassium channel domain. Structure 15:235–244.
[41] Jogini V, Roux B. 2007. Dynamics of the Kv1.2 voltage-gated K^+ channel in a membrane environment. Biophys J 93:3070–3082.
[42] Bond PJ, Sansom MSP. 2006. Insertion and assembly of membrane proteins via simulation. J Amer Chem Soc 128:2697–2704.
[43] Lindahl E, Hess B, van der Spoel D. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Molec Model 7:306–317.
[44] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690.
[45] Granseth E, von Heijne G, Elofsson A. 2005. A study of the membrane-water interface region of membrane proteins. J Mol Biol 346:377–385.
[46] Long SB, Tao X, Campbell EB, MacKinnon R. 2007. Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment. Nature 450:376–382.
[47] Periole X, Huber T, Marrink SJ, Sakmar TP. 2007. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. J Am Chem Soc 129:10126–10132.
[48] Yeagle PL, Bennett M, Lemaitre V, Watts A. 2007. Transmembrane helices of membrane proteins may flex to satisfy hydrophobic mismatch. Biochim Biophys Acta 1768:530–537.
[49] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80:505–515.
[50] Hilf RJC, Dutzler R. 2008. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature 452:375–379.

This paper was first published online on iFirst on 21 October 2008.

Downloaded By: [University of Oxford] At: 15:53 27 November 2008